THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Kirk J. Johnstone

TITLE OF THESIS An Application of Two Markov Chain Models to Precipitation at Some Alberta Locations

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Science

YEAR THIS DEGREE GRANTED 1980

THE UNIVERSITY OF ALBERTA

AN APPLICATION OF TWO MARKOV CHAIN MODELS TO PRECIPITATION AT SOME ALBERTA LOCATIONS

by

© KIRK J. JOHNSTONE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

IN

METEOROLOGY

DEPARTMENT OF GEOGRAPHY

EDMONTON, ALBERTA

FALL, 1980

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "An Application of Two Markov Chain Models to Precipitation at Some Alberta Locations", submitted by Kirk J. Johnstone in partial fulfilment of the requirements for the degree of Master of Science in Meteorology.

Dedication

For Mary-Marguerite,
whose love and patience
has seen us through this
work during our first
two years.

Abstract

Two Markov chain models, proposed recently by Katz and by Todorovic and Woolhiser for daily precipitation, were applied to data from Beaverlodge, Edmonton, and Medicine Hat, Alberta. Model parameters were estimated from a development sample and the resulting distributions were compared with independent data samples. The distributions calculated for the number of wet days during the month are nearly the same for both models and represent adequately the precipitation occurrence process. The distributions calculated for the total monthly precipitation are also adequate. The models do not represent adequately the climatology of the maximum daily precipitation. This shortcoming is attributed primarily to the inability of the Gamma distribution to represent correctly the daily precipitation amounts.

A few of the assumptions required by the development of the models were examined. These included the assumption that the precipitation process is stationary, that the occurrence of precipitation is a first-order Markov chain, that the precipitation amounts on consecutive wet days are independent, that the precipitation amounts are dependent on the wet-dry state of the previous day, and that the total monthly precipitation is independent of the number of wet days during the month. The simple techniques that were used did not offer any conclusive evidence that the precipitation

process is nonstationary. Two selection critera, proposed recently by Akaike and Schwartz, were used to show that a first-order Markov chain is appropriate for the cases considered. Correlation analysis showed that consecutive wet-day amounts are independent in most cases, but graphical displays indicated that there is some functional dependence. A maximum likelihood test showed that the distributions of precipitation amount after a wet or dry day are significantly different for Beaverlodge only. Correlation analysis provided conclusive evidence that the total monthly precipitation is dependent upon the number of wet days during the month.

An unexpected result is the large sampling fluctuation exhibited by the precipitation characteristics that were examined. Because of this it is not possible to accept or reject conclusively the models' representation of the ensemble of precipitation time series.

Acknowledgments

I would like to thank Dr. Hage for his assistance during the course of this work.

I also wish to thank Dr. McLeish and Dr. Lozowski for taking the time to serve on my examining committee.

I wish to express my gratitude to Laura Smith for her advice and for typing some of the equations.

I wish to thank the Atmospheric Environment Service, Environment Canada, for the prompt provision of the data used in this study.

This study was conducted while on Education Leave from the Atmospheric Environment Service, Environment Canada.

Table of Contents

List of Tables

List of Figures

Figure		Page

Figure Page

Figure		Page

CHAPTER 1.

Introduction

1.1 The Nature and Importance of Precipitation

Even the most optimistic weather forecasters or researchers cannot expect to predict the daily occurrence of precipitation for time periods of more than a few days. The daily precipitation at a specified location has a stochastic nature; the precipitation amount is a random variable, and the record of daily precipitation is a precipitation time series. The precipitation record available for a location is one realization of the ensemble of series that is possible. To understand the probabilistic nature of the daily precipitation that occurs during long time periods it is worthwhile to analyze the existing records, and attempt to model the ensemble of daily precipitation time series.

Precipitation is important to the general public. Aside from being a common opening topic of conversation, precipitation, or the weather in general, affects many social and economic activities, particularly during long periods of dry, wet, hot, or cold weather. The dry spring of the Prairie provinces and the extremely hot summer of the American mid-west during 1980 are two recent examples whose effects received almost daily attention by the news media. The lengths of dry or wet spells and weather cycles motivated the earliest studies of precipitation records by such authors as Besson (1924), Weiss (1944), Jorgensen (1949),

Longley (1953), and Gabriel and Neumann (1962).

To utilize a hydrological resource its full potential must be known (Meheriuk, 1972). Agriculture and engineering require knowledge of potential precipitation totals during specified time periods, and so the studies and modeling of precipitation were extended to include the maximum daily and total precipitation that could be expected during a specified time period.

The stochastic modeling of precipitation will remain an area of active research because precipitation is closely linked to all hydrological applications. Precipitation and streamflow models provide important input to engineering design (Haan, 1977), pollution control, forestry and agriculture (Farmer and Homeyer, 1974), and the provision of hydro-electric power. Many meteorologists can use information about precipitation time series. Modeled precipitation distributions provide a more detailed climatological reference than the commonly used means and extremes. The modeled reference could also be used as a standard to judge the skill of long range forecast techniques.

1.2 Past Studies and the Present Models

The models used in this study are based on persistence. The existence of persistence in meteorological variables is well known (Besson, 1924; Weiss, 1944; Jorgensen, 1949; Brooks and Carruthers, 1953). In particular, Hannan (1955)

discussed the "lack of independence between rainfalls on days near each other in time." Brooks and Carruthers (1953) first suggested that a simple Markov chain could be used to describe the persistence of daily precipitation. Gabriel and Neumann (1962) successfully applied Markov chain theory to records of daily rainfall at Tel Aviv.

The concept that the probability of the occurrence of rain today depends on whether or not rain occurred yesterday, but not on whether or not rain occurred two days ago is the basis of a first order Markov chain, often called a simple Markov chain.

Subsequent to Gabriel and Neumann's first application of the Markov chain approach, its use became commonplace. Caskey (1963) found that theoretical probabilities calculated using Markov chain theory agreed closely with empirical values for the probability of precipitation occurrence at Denver, Colorado (Topil, 1963). Hopkins and Robillard (1964) used a simple Markov chain, with some success, to model the occurrence of precipitation for summer months at three locations in the prairie provinces. Weiss (1964), and Feyerherm and Bark (1965, 1967) also found that the simple Markov chain adequately represented the probability distribution for the occurrence of precipitation. However, the use of the simple Markov chain did not have unlimited success; Wiser (1965), Lowry and Guthrie (1968), and Green (1970) found it necessary to propose more general models. In some cases a higher-order Markov chain could match the

results of the more general models.

To meet the demands of engineers, Todorovic and Yevyevich (1967) proposed a stochastic model that was capable of providing information on the amounts of precipitation that could be expected. The development of their model required the assumption that occurrences of precipitation are serially independent. Verschuren (1968) applied the model to records for two locations in the United States, and Meheriuk (1972) applied the model to precipitation records for a number of locations in Alberta. Meheriuk noted that the assumption of independence was a shortcoming of that model.

Further work by Todorovic and Woolhiser (1974, 1975) resulted in a model that was capable of calculating the probability distribution for the maximum daily and total amount of precipitation occurring during an n-day period. They used a simple Markov chain to model the persistence of the occurrence of precipitation. The Todorovic and Woolhiser (TW) model is one of the two examined here.

As an alternative, Katz (1974) pointed out that a recurrence relation for a simple Markov chain could be used to calculate the probability distribution for the number of wet days during a given period. Later, Katz (1977a) proposed an extension of the recurrence relation approach that was capable of calculating the distributions for the maximum daily and total amount of precipitation in an n-day period. Katz's recurrence relation model is the second examined

here.

Both of the models examined in this study used a simple Markov chain to model the daily occurrence of precipitation. Given that precipitation had occurred, a well-known distribution function was then used to determine the amount of precipitation that had occurred.

This approach to the modeling of precipitation (Thom 1951, 1968), rather than simply fitting the observed data to a well-known distribution function will, I hope, provide a better approximation to the ensemble of possible time series. The approach is justified because the meteorological systems causing measurable precipitation are different than those causing no precipitation. The use of theoretically derived distributions permits a better understanding of the circumstances under which the distributions are reasonable approximations to those observed than does simply fitting the observed data to a well known distribution function (Meheriuk, 1972). There is also a better chance of explanation and correction of differences between the calculated and observed distributions than there is when the latter procedure is used.

Beyond the common use of a Markov chain to model the daily occurrence of precipitation, the approaches of Todorovic-Woolhiser and Katz diverged. The Todorovic-Woolhiser approach was to obtain an exact solution to the stochastic problem. Katz used an iterative computational technique which allowed a more general model. For example,

the Katz approach allowed nonstationarity in the Markov chain transition probabilities and the use of a number of statistical distributions to describe daily precipitation totals. The TW model required stationary transition probabilities and the exponential distribution was used to describe daily precipitation totals.

The theoretical modeling of precipitation has been done using many models, including multi-state Markov chains (Haan et. al., 1976; Selvalingam and Miura, 1978), regional models (Richardson, 1977), the models used in this study, and others.

1.3 A Few Problems

The stationarity and homogeneity of the data available for parameter estimation were of crucial importance to the stochastic models (Yevyevich, 1972; Haan, 1977). Simple statistical techniques were applied to the data to attempt to identify nonstationarity or inhomogeneity. The results were cautiously interpreted because secular trends and long term periodicities are controversial topics. Statistically significant results can be artificially caused by the measurement or analysis of the data. The use of a historical summary of the observing site and procedures, as recommended (Yevyevich, 1972), was used to try to identify physical causes for statistically significant changes in the record.

A major test of the models was whether or not the modeled distributions satisfactorily reproduced a

distribution from an independent data sample. Despite the constant development and testing of precipitation models, there is little literature on the comparison of model results with independent data sets. Often, the entire record is used for parameter estimation; statistics calculated by Monte Carlo simulations are then compared with statistics of the development data to judge a model's ability. The critical test of a model, a comparison of the model's results with an independent data sample, is often not done.

Klemeš and Bulu (1979) used such a test to determine the capabilities of three stochastic hydrologic models to represent the ensemble of monthly steamflow values for the Elbe river. The title of their paper, "Limited Confidence in Confidence Limits Derived by Operational Stochastic Hydrologic Models," summarized their conclusions. Part of the present study was to compare the modeled distributions with independent distributions and show the sampling fluctuation of observed probability distributions.

1.4 Study Objectives

The first objective of the present study was to determine the abilities of the TW and Katz models to represent the climatological probability distributions for the number of wet days, the maximum daily precipitation, and the total precipitation amount during an n-day period. The second objective was to examine the many assumptions required by

the modeling of a climate record, and by the models. Testing of the assumptions may enable the criticism and improvement of past and present models. The third objective was to briefly examine how representative an observed precipitation record is of the ensemble of precipitation series, and to attempt to answer concomitant questions about the reliability of model results. The fourth objective was to gain experience in the application of statistical techniques by examination of a data set. It was not the purpose of this study to choose an operational model for Alberta.

CHAPTER 2.

The Data

2.1 Data Selection

The nature of this study made the selection of a number of data sets necessary. To meet the objectives of the study, long and complete daily climatological records were required. This was to ensure that adequate data were available for parameter estimation and model testing. Because site changes can introduce discontinuities into a climatological record, the data sets considered for selection were required to have a readily available historical record. Ideally, the history would show the sites had not been moved since observations began. Unfortunately, the longer climatological records are a series of observations taken at a number of station locations. Accordingly, stations whose sites had been located in a small area were sought.

The Climatological Station Data Catalogue (1976) was used initially to select a number of data sets for the study. The initial selection was reduced to six sets of data for stations in Alberta on the basis of the historical summaries given by Lachapelle (1977). The complete daily climatological record for each of the six locations selected was then obtained from the Atmospheric Environment Service (AES).

The Beaverlodge CDA (Canadian Department of Agriculture) record was selected for a preliminary investigation

because the historical record showed that it had a single, smallest, and possibly least significant change in site location. The conclusions were that the sixty-six years of record available, 1913 to 1978, were insufficient for parameter estimation and testing of the model. Confidence limits, based on the Kolmogorov-Smirnov test statistic at the five percent level of significance, were found to be so large that nearly any theoretical distribution calculated would be accepted as the same as the distribution obtained from the independent test data when only a small sample of test data were available, as in the Beaverlodge case. Consequently, the two other stations selected for study were those having the longest periods of record: Edmonton (1880-1978) and Medicine Hat (1883-1978).

The climatological records of Beaverlodge, Edmonton, and Medicine Hat were visually examined for missing months of data. The records were generally found to be missing complete months of record in the station's first few years of operation only.

The climatological records used were split into development and independent data samples. Forty-five to fifty years of data were desired for parameter estimation; consecutive years were used for ease of processing.

The Beaverlodge record had been split into a development sample consisting of the years 1914 to 1958 and a test sample running from 1959 to 1978. The Edmonton development sample was chosen to be from 1883 to 1932; the test data

from 1933 to 1978. The Medicine Hat record was split into periods of 1884 to 1933 and 1934 to 1978, the first being the development data and the second the test data.

2.2 Site Histories

Unfortunately, the use of stations with the longer periods of record sacrificed the third selection criterion to an extent. Both the Medicine Hat and Edmonton site locations have changed a number of times during the stations' periods of record. Details of the site changes are available in Lachapelle's (1977) thesis; for the purpose of this study a brief summary is all that is required.

The Edmonton site has been moved five times since the first observations were taken in 1880 at Fort Edmonton, then located on what are now the grounds of the legislature. The first move, in 1882, to just north of Jasper Avenue on what is now 115 Street cannot affect this study because data obtained prior to 1882 were not used. The second site was 3km south of the site's present location at the Edmonton Municipal Airport. A second move in April 1912 was to 63rd Street in the Highlands, approximately 5.2km east of the site's present location. Observations were taken at that location until 1942 when the station was closed.

In September 1937 a new station was opened at the Edmonton Municipal Airport. A number of site changes on the airport grounds have occurred since that time, but these were minor relocations and unimportant. The present site is

at 53° 35′ North, 113° 30′ West at an elevation of 670.5m.

The record from the Highlands site was combined with that from the airport at the end of 1937. The record combination can be considered a site change. With the site transition, observations became the responsibility of a trained observer at the airport. Consequently, the end of 1937 was a time when a discontinuity could have been introduced into the record, because of both the site change and possible changes in observation procedure.

In 1883, the first complete month of observations was taken in the town of Medicine Hat. A number of site changes, all to new sites within the South Saskatchewan River Valley, occurred during the period 1883 to 1930. The most significant site changes were made in 1930 and 1931 when the site was moved out of the river valley, and 4.3Km across town to what is now the airport. The site has remained out of the valley since 1931. The present location is 50° 01′ North, 110° 43′ West, at an elevation of 720.8m.

The Beaverlodge site has been moved once during the station's history. The second site is 373.7m south of the first, in slightly rolling terrain. The site was moved on 1 January 1958 after a three year comparative study of observations taken at both sites. The study did not reveal any significant difference in the precipitation records (Carder, 1962). The present Beaverlodge site is located in a field at 55° 12′ North, 119° 25′ West, west of Grande Prairie, Alberta, at an elevation of 731.5m. The station

locations are shown in Figure 1.

The climatological records for each station were examined in an attempt to determine if the site changes (Beaverlodge, 1957; Edmonton, 1937; Medicine Hat, 1931) had introduced a discontinuity into the data. The techniques used are discussed in Chapter 5.

2.3 Observations

The complete climatological records for the three stations investigated were supplied on magnetic tape by the AES. The total precipitation records were then sorted and transferred to another magnetic tape before further processing.

Each station's precipitation record consisted of a series of monthly computer records, each of which included the station identifier, year and month of the observations, and up to thirty-one daily observations of precipitation amount with flags. The daily observations of total precipitation were in tenths of a millimetre. The flags gave information about the daily values, for example, an M indicated that the observation was missing while an E meant that the amount was estimated.

The precipitation amount was defined (MANOBS, 1971, 1976) to be the vertical depth of water which reaches the ground in the stated period, i.e., one day. The total precipitation recorded is the rainfall, water equivalent of snowfall, or sum of the two which reaches the ground in a

twenty-four hour period. Rainfall totals were determined by measuring in a graduated cylinder the water catch of a copper gauge with a mouth of 25.4cm² that was exposed 30.5cm above level ground. In the 1970's the copper gauges were replaced by the Type B gauge, a plastic gauge with a mouth of 25.4cm² that was exposed 40.6cm above level ground. The old gauges were replaced by the new, larger capacity, gauges to eliminate loss of data because of overflow.

Replacement of the copper gauges was a possible source of discontinuity in the record, but the change caused by a 10cm increase in the height of the gauge mouth is probably insignificant compared to catch changes because of site movements, or catch errors resulting from other causes, e.g., wind. A study of catch changes or errors because of equipment changes was beyond the scope of this work, and in any event, only the last eight years of the records would have been affected by the change in rain gauges.

For many years the water equivalent of snowfall was obtained by first averaging a number of ruler snow depth measurements with allowances made for drifting. The water equivalent was assumed to be ten percent of the average depth of the new fallen snow. During and after 1960, snow gauges were introduced at principal observing sites to measure the actual amount of the water equivalent of freshly fallen snow. The introduction of snow gauges was another possible source of discontinuity in the precipitation record that was not examined.

In Canada, prior to 1976, precipitation amounts were measured and recorded in inches. After 1975, precipitation amounts were measured and recorded in millimetres. The precipitation records used were computer processed by the AES; the measurements in hundredths of an inch were converted and rounded to tenths of a millimetre for the period of record to 1976.

The difficulty of measuring small amounts of rain or snowfall has resulted in the concept of a trace of precipitation. Prior to 1976, a precipitation amount less than five one-thousandths of an inch was recorded as a trace (flag T) and zero amount was recorded. The smallest recorded amount was one one-hundredth of an inch. Since 1976, less than one-tenth of a millimetre was recorded as a trace and the smallest recorded amount was two-tenths of a millimetre.

Measurable precipitation is a precipitation amount greater than a trace. For the purposes of this study a wet day was defined to be a day for which a measurable amount of precipitation was recorded. A dry day was a non-wet day.

Strictly speaking, day meant climatological day. Meheriuk (1972) outlined the history of the climatological day for first order and ordinary climate stations. According to Meheriuk, the stations selected for this study were first order--the order determined by the elements observed and the number of observations each day--with Beaverlodge considered a climate station prior to 1935 and after 1955.

The relevant points of Meheriuk's summary are given here.

For first order stations only one observation was taken each day, at 0700 LST, during the years 1878 to 31 May 1924. The climatological day began on day t following the 0700 LST observation and ended on day t+1 at 0700 LST; all observed elements were credited to day t+1.

From 1 June 1924 to 31 December 1932, for stations taking observations at 0700 and 1900 LST each day, the climatological day began on day t following the 1900 LST observation and ended at 1900 LST on day t+1. Observed elements were credited to day t+1. Stations taking one observation each day used the 1878 to 1924 procedure.

The climatological day was brought into line with the usual notion of a day on 1 January 1933. Stations taking one observation each day did so at 0630 LST. The climatological day began after the 0630 LST observation on day t and ended at the 0630 LST observation on day t+1; observed elements were credited to day t. Some stations took a second observation at 1830 LST, but the climatological day was the same as that for stations taking only one observation.

Another change was made on 1 January 1941. Most stations were then required to take four observations each day, at 0130, 0730, 1330, and 1930 GMT. Some stations took from one to three observations only. Elements observed during the climatological day, from 0730 GMT on day t to 0730 GMT on day t+1, were credited to day t.

On 1 January 1955 the observation times were moved one hour earlier and the climatological day ran from after the 1230 GMT observation on day t through to the 1230 GMT observation on day t+1. The observation times were moved another 0030 hour earlier, with the same shift in the climatological day, on 1 June 1957.

From 1 July 1961 to the present, most stations were required to take four observations each day, some taking from one to three. Observation times were 0000, 0600, 1200, and 1800 GMT. The climatological day began on day t following the 0600 GMT observation and ended on day t+1 at 0600 GMT. Observed elements were credited to day t.

Observers at ordinary climate stations (Beaverlodge, with the exception of the period 1935 to 1955) were encouraged to take observations twice a day, as close to 0800 and 1700 LST as possible. Since 1933 the climatological day has begun after the 0800 LST observation on day t and ended with the 0800 LST observation on day t+1. Observed elements were credited to day t.

Changes in the climatological day are other possible sources of discontinuities in the recorded data. For example, prior to 1933, if measurable precipitation occurred after 0700 LST on day t, at a station taking one observation per day, day t+1 was recorded as wet. After 1 January 1933, day t would be recorded as wet. The change in procedure may be responsible for a slight error in estimation of the initial and transition probabilities for the Markov chain,

but the error should be small since a large data sample was used for parameter estimation. Also, changes in the observation time may have caused a discontinuity in the precipitation amounts recorded. The discontinuity would be small and detection of such a discontinuity was beyond the scope of this work.

2.4 Data Errors

Errors in the data are problems that are difficult to contend with. No attempt was made to identify faulty observations because a quality check of observations is routinely carried out by the AES. The data were accepted to be accurate despite the possibility of erroneous values in the record. But errors are important in a study such as this, particularly when the results are applied, so a brief discussion on errors is included.

Haan (1977) identified three general sources of error in hydrologic data. They are: measurement error, data transmission error, and processing error. A discussion of these error sources for a precipitation record follows.

Measurement error in precipitation data may have a number of causes. First, improper exposure of a rain gauge may result in water blowing off trees or buildings into the gauge. To alleviate this problem gauges are supposed to be located, whenever possible, on level ground with no obstacles nearer than four times their vertical height (Canadian Normals, Precipitation, 1973). Second, there is a loss

because of wetting the receiver when transferring the precipitation to the measuring cylinder. The loss due to wetting should be approximately 0.25mm or less (Meheriuk, 1972). Third, a loss to evaporation becomes significant when observations are taken up to twenty-four hours apart. According to Meheriuk, Weisner (1970) claimed that the mean error, in Russian investigations, caused by evaporation is from three to five percent of the annual total precipitation, or 0.51mm or less for individual measurements. Fourth, the maximum error in precipitation measurement is caused by the wind. Wind causes a deficiency in the rain catch, and in the snow catch at low speeds (Meheriuk, 1972). For higher wind speeds snow also blows into the gauge, causing an excessive measurement. For unshielded gauges, errors in the rain catch can amount to twenty and fifty percent at wind speeds of ten and forty knots. Snow catch errors can amount to forty and seventy percent for the same wind speeds (Meheriuk, 1972). Shielding the gauges reduces the error caused by the wind. Fifth, inadvertent bias by the observer can cause errors. A bias towards nice numbers, e.g., multiples of one-tenth of an inch, has possibly been introduced into the records for some sites. This bias will be commented on further in Chapter 4.

Data transmission errors can result from illegible writing, mistakes in card punching, or vague explanation of observations because of coding methods. Little can be done about these error sources, the author hopes such errors are

infrequent.

Explanations about daily observations were provided by the daily flags included in the record. The flags used for total precipitation were:

1. A...accumulated amount, previous value C or L,
2. C...precipitation occurred, amount uncertain, recorded value zero,
3. E...amount estimated,
4. F...amount accumulated and estimated,
5. L...precipitation may or may not have occurred, recorded value zero,
6. M...missing, and
7. T...trace amount, recorded value zero.

In this study a recorded amount greater than zero or a flag C was taken to be a wet day. Consequently, if a C day preceded an A day in the record they were both considered wet. As Meheriuk (1972) pointed out, there is no way of determining if precipitation occurred on the A day or not. The presence of A's, C's, F's, or L's in the record can cause errors in the determination of the model parameters. Such errors were not significant in this study because there were few A's, C's, F's, or L's in the records used.

A processing error has been introduced into the data taken prior to 1976 by its conversion to metric values. However, the roundoff of precipitation values to tenths of a millimetre in the conversion is a small error, and it is not important to this study.

CHAPTER 3.

The Theory

3.1 Definition of the Markov Chain and Precipitation Process

The simple Markov chain is defined to be a sequence of discrete random variables, $\{Y_t; t=1,2,...\}$, with the property that the conditional distribution of Y_t, given Y_{t-1}, $Y_{t-2}, ...$ depends on Y_{t-1}, but not on Y_{t-2}, $Y_{t-3}, ...$ Let the S+1 discrete values or states which the Y_t's assume be denoted by 0,1,2,...,S. The simple Markov process is characterized by the property:

$$Pr(Y_t = j | Y_{t-1} = i,\ Y_{t-2} = l, ...) = Pr(Y_t = j | Y_{t-1} = i),$$
$$i, j, l, ... = 0, 1, 2, ... S.$$

The probability $Pr(Y_t = j | Y_{t-1} = i)$ is called the transition probability p_{ij}; $i,j=0,1,...S$, and represents the probability of a transition from state i to state j. A brief discussion of Markov chain terminology is given in Appendix A, or is available in many texts (Feller, 1957; Cox and Miller, 1965).

Daily precipitation is a bivariate stochastic process represented by: $\{(Y_t, X_t); t=0,1,2,...\}$.

$$\text{Let } Y_t = \begin{cases} 0, & \text{if the t-th day is dry,} \\ 1, & \text{if the t-th day is wet;} \end{cases}$$

then the sequence $\{Y_t; t=1,2,...\}$ represents the stochastic occurrence of precipitation. The amount of precipitation that occurs on the t-th day is denoted X_t; note that $X_k = 0$ for $Y_k = 0$.

The assumptions about the X_t process used by the two models that are considered in this study are different and are given later. Both models assume the Y_t process to be a simple two-state Markov chain.

The method of calculating the density function for the number of wet days, during an n-day period, is given in the next section. First, Gabriel's (1959) derivation and results, used by Todorovic and Woolhiser, are presented. This is followed by the recurrence relation which was suggested by Katz.

3.2 Number of Wet Days in an n-Day Period

The following derivation is after Gabriel (1959), who developed an expression for the number of successes in n dependent trials.

A sequence of n days represents n trials $Y_1, Y_2, \ldots, Y_n$ following an initial trial Y_0. Let the stochastic variable representing the wet-dry state of day t, Y_t, be a simple two-state Markov chain. Then $p=Pr(Y_0=1)$ is the initial probability of a wet day, i.e., a success. The probability of a dry-to-wet transition is $p_{01}=Pr(Y_t=1|Y_{t-1}=0)$ and $p_{11}=Pr(Y_t=1|Y_{t-1}=1)$ is the probability of a wet-to-wet transition. Assume p, p_{01}, and p_{11} are independent of t during the n days.

The number of wet days in the n days or trials is $s=\sum_{t=1}^{n} Y_t$. Feller (1957) has shown that s is asymptotically normally distributed with an expected value

$$E(s) \sim np$$

and

$$Var(s) \sim np(1-p)[(1+d)/(1-d)],$$

where $d = p_{11} - p_{01}$. But this result gives neither the exact distribution for small n nor the rapidity of approach to normality (Gabriel, 1959). The distribution for small n was obtained with the following argument.

When s wet days occur in n days there will be a number of changes from a wet day (including day t=0) to a dry state on the next day, and vice versa. Denote the number of changes by C. Define "a" to be the least integer not smaller than (1/2)(C-1) and "b" to be the least integer not smaller than C/2.

Consider the case of an initial wet day. Then s wet days with C changes will involve b wet-to-dry transitions and "a" dry-to-wet transitions. Of the changes, b must be wet-to-dry transitions, otherwise it is not possible to have C changes when the initial day is wet. The remaining "a" changes must be dry-to-wet transitions. Since there are "a" wet days during the n days resulting from dry-wet transitions an additional s-a wet days must occur as the result of wet-to-wet transitions. Similarly n-s-b dry-to-dry transitions occur. The probability of any one arrangement of s wet days with C changes in n days is

$$(1-p_{11})^{b}(p_{01})^{a}(p_{11})^{s-a}(1-p_{01})^{n-s-b}$$

or

$$p_{11}^{s}(1-p_{01})^{n-s}(p_{01}/p_{11})^{a}[(1-p_{11})/(1-p_{01})]^{b}.$$

Any arrangement of n days with s wet days and C changes involves "a" dry-wet transitions which may occur before any "a" of the s successes, in any of $\binom{s}{a}$ different positions. Also, b changes occur before dry days, of which the first must occur before the first dry day and the rest can be arranged in $\binom{n-s-1}{b-1}$ different ways. Given arrangements of both kinds of changes the total number of possible arrangements of C changes among n trials with s wet days is

$$\binom{s}{a}\binom{n-s-1}{b-1}$$

Hence, the probability of s wet days with C transitions in an n-day period following an initial wet day is

$$\Pr(s,c|n,Y_o=1) = \binom{s}{a}\binom{n-s-1}{b-1} p_{11}^{\ s} (1 - p_{01})^{n-s} \left(\frac{1 - p_{11}}{1 - p_{01}}\right)^b \left(\frac{p_{01}}{p_{11}}\right)^a .$$

The number of changes C may be any positive integer up to $C_1 = n+1/2-|2s+1/2-n|$. Thus the probability of s wet days in an n-day period following an initial wet day is obtained by summing over all possible values of C and is

$$W_1(s;n) = p_{11}^{\ s} (1 - p_{01})^{n-s} \sum_{C=1}^{C_1} \binom{s}{a}\binom{n-s-1}{b-1}\left(\frac{1 - p_{11}}{1 - p_{01}}\right)^b \left(\frac{p_{01}}{p_{11}}\right)^a$$

For an initial dry day, b of the transitions must be dry-to-wet transitions. The remaining "a" changes are wet-to-dry transitions. A similar argument shows the probability of s wet days in an n-day period following an initial dry day to be

$$W_o(s;n) = p_{11}{}^s (1 - p_{01})^{n-s} \sum_{C=1}^{C_o} \binom{s-1}{b-1}\binom{n-s}{a}\left(\frac{1 - p_{11}}{1 - p_{01}}\right)^a \left(\frac{p_{01}}{p_{11}}\right)^b ,$$

where C_0 is $n+1/2-|2s-1/2-n|$. When d=0, both probabilities become that of the binomial distribution.

The probability of s wet days among the next n days is given by:

$$W(s;n)= Pr(s;n)= pW_1(s;n) + (1-p)W_0(s;n). \qquad (2.1)$$

Gabriel and Neumann (1962) admitted that calculating the required probabilities by hand is a tedious chore.

Katz's (1974) approach is simpler. Katz utilized a recurrence relation, earlier derived by Helgert (1970), for $W_0(s;n)$ and $W_1(s;n)$. Noting that (2.1) for $W(s;n)$ was arrived at by conditioning on whether the initial day was wet or dry ($Y_0=1$, or $Y_0=0$), recurrence relations for $W_0(s;n)$ and $W_1(s;n)$ were obtained by conditioning on Y_1.

In order to have s wet days in n days either:

1. $Y_1=0$, and there are s wet days in the n-1 remaining days, or
2. $Y_1=1$, and there are s-1 wet days in the n-1 remaining days.

The first case cannot occur if s=n, and the second is impossible for s=0.

Given that $Y_0=0$, the first case occurs with probability $p_{00}W_0(s;n-1)$, and the second with probability $p_{01}W_1(s-1,n-1)$. Similarly, for $Y_1=1$, the first case occurs with probability $p_{10}W_0(s;n-1)$, and the second with

probability $p_{11}W_1(s-1;n-1)$. Hence,

$$W_0(s;n) = p_{00}W_0(s;n-1) + p_{01}W_1(s-1;n-1)$$

and

$$W_1(s;n) = p_{10}W_0(s;n-1) + p_{11}W_1(s-1;n-1)$$

for $s=0,1,2,...,n$ and $n=1,2,....$ Initial conditions are simply that the probability of zero wet days in zero days is one, for the initial day either wet or dry, i.e.,

$$W_0(0;0) = W_1(0;0) = 1.$$

The constraints are formulated as

$$W_0(n;n-1) = W_1(-1;n-1) = 0.$$

Given the transition probabilities p_{ij} and the initial probability p, $W_0(s;n)$ and $W_1(s;n)$ can be computed recursively for $s=0,1,2,...n$ and $n=1,2,....$ Eq. (2.1) then gives the distribution of the number of wet days in an n-day period.

The recurrence relation method of calculation leads to a natural introduction of time-dependent transition probabilities. Consequently, the Katz model was programmed to allow for nonstationarity of the transition probabilities.

The difficulty of writing a computer routine for Gabriel's exact approach has been overcome by Todorovic and Woolhiser (1974). A version of Todorovic and Woolhiser's (TW) routine is given in Appendix B. The original routine given by Todorovic and Woolhiser was rewritten for implementation in this study.

Gabriel's exact approach requires a homogeneous Markov chain, so constant transition probabilities were used in the

TW model.

3.3 The TW Model

The sequence $\{X_t; t=1,2,...,n\}$ represents the amount of precipitation occurring on the n days. Renumbering the X_t's so that $X_k = X_t$, $k=0,1,2,...,s$ and $t=0,1,2,...,n$, the X_k's denote the amount of precipitation on the kth wet day. The kth wet day may be any day after the (k-1)th wet day.

The largest daily value of precipitation in the n-day period, M_n, is given by:

$$M_n = \max X_k,\ 0 \le k \le s, \text{ where } s = \sum_{j=1}^{n} Y_j .$$

The total amount of precipitation in the n-day period, T_n, is given by:

$$T_n = \sum_{k=0}^{s} X_k,\ X_0 = 0,\ s = \sum_{j-1}^{n} Y_j .$$

That $Pr(M_n = 0)$, and $Pr(T_n = 0)$ both equal $Pr(Y_1 = 0, ..., Y_n = 0)$ immediately follows.

The sequence of events $\{s=0\}$, $\{s=1\}, ...,$ $\{s=n\}$, by definition, represents a finite partition of the sample space. This partition means that

$$\{s=i\} \cap \{s=j\} = \theta \text{ for } i \ne j, \tag{2.2}$$

and $\sum_{i-1}^{n} Pr(s=i) = 1$, where θ is the null set.

3.3.1 Maximum Daily Precipitation

The distribution function $G(x)$ for the maximum daily precipitation amount during n days, M_n, is defined:

$$G(x)= Pr(M_n \le x), \quad x \ge 0.$$

The finite partition of s means

$$G(x)=Pr(M_n \le x, \bigcup_{j-0}^{n} \{s=j\}),$$

so

$$G(x)=\sum_{j-0}^{n} Pr(\max X_k \le x, s=j); \quad 0 \le k \le s.$$

Assuming the $X = X_1, X_2, ..., X_s$ are independent of s

$$G(x)=\sum_{j-0}^{n} Pr(\max X_k \le x) Pr(s=j), \quad 0 \le k \le j.$$

The further assumption that $X_1, X_2, ..., X_s$ are independent, identically distributed random variables, such that $V(x)= Pr(X_k \le x)$, leads to

$$G(x)= Pr(s=0)+\sum_{j-1}^{n} (V(x))^j Pr(s=j), \tag{2.3}$$

because

$$Pr(\max X_k \le x)=\prod_{k-1}^{j} Pr(X_k \le x)= (V(x))^j, \quad 1 \le k \le j.$$

The distribution G(x), for the maximum daily precipitation amount during n days M_n, can be numerically calculated using (2.3) if V(x) and Pr(s=j) are known.

3.3.2 Total Precipitation Amount

We define $H_n(x)$ to be the distribution function of T_n, so that

$$H_n(x)=Pr(T_n \le x), \quad x \ge 0.$$

Then on the basis of (2.2),

$$H_n(x)=Pr(\sum_{k-0}^{s} X_k \le x, \bigcup_{j-0}^{n} \{s=j\})=\sum_{j-0}^{n} Pr(\sum_{k-0}^{j} X_k \le x, s=j).$$

That

$$H_n(x)=\sum_{j=0}^{n} Pr(\sum_{k=0}^{j} X_k \leq x)Pr(s=j) \tag{2.4}$$

follows, because $T_j = \sum_{k=0}^{j} X_k$ is assumed independent of j. Rewriting (2.4) gives

$$H_n(x)= \sum_{j=0}^{n} Pr(T_j \leq x)Pr(s=j)$$

$$=Pr(s=0)+\sum_{j=1}^{n} Pr(T_j \leq x)Pr(s=j) \tag{2.5}$$

where $X_0 = 0$. The result (2.5) can be used to calculate $Pr(T_n \leq x)=H_n(x)$, if $Pr(s=j)$ and $Pr(T_j \leq x)$ are known.

3.3.3 Application

The probability that s is equal to j, $Pr(s=j)$, is simply the probability of exactly j wet days in an n-day period. This probability can be calculated using either the exact method of Gabriel, or the recurrence relation approach of Katz. Todorovic and Woolhiser (1974) used the first approach, it was used also in the present TW model.

The distribution $V(x)$ for the X_k, assumed independent and identically distributed, must be provided. The distribution can be selected from either an observed or theoretical distribution. The use of an observed $V(x)$ would require the tabulation of distribution values for discrete x. Using a theoretical distribution, with parameters estimated from the observed data, is a more common approach.

The probability that the sum of the random variable X_k is less than or equal to x, $Pr(T_j \leq x)$, must also be determined. In order to be consistent within the model, and

following Todorovic and Woolhiser, the X_k are assumed to be distributed according to a special case of the gamma distribution, the exponential:

$$V(x)=\begin{cases}1-\exp(-\lambda x), & x\geq 0,\\ 0, & x<0,\end{cases} \tag{2.6}$$

where λ is a scale parameter. In this case, Todorovic and Woolhiser (1974) show that

$$\Pr(T_j \leq x)=(\lambda^j/\Gamma(j))\int_0^x u^{j-1}\exp(-\lambda u)du, \tag{2.7}$$

the gamma distribution with shape parameter j, and scale parameter λ.

A well known theorem of statistics (Kendall and Stuart, 1963) states that the characteristic function,

$$\phi(u)=\int_{-\infty}^{\infty}\exp(iux)dF,$$

where F(x) is the cumulative distribution for X, uniquely determines the distribution function. Following Todorovic and Woolhiser (1974), the characteristic function for the total amount of precipitation is

$$\phi(u)=\int_{-\infty}^{\infty}\exp(iuT_j)dH=E[\exp(iuT_j)]$$

where E represents the expectation of the bracketed value, H is the distribution function for the total amount of precipitation in j days, and u is a characteristic function parameter. Now,

$$E[\exp(iuT_j)]=E[\exp(iu\sum_{k-0}^{j}X_k)]=$$

$$E[\exp(iuX_1)\exp(iuX_2)\ldots\exp(iuX_j)]=$$

$$E[\exp(iuX_1)]\ E[\exp(iuX_2)]\ldots E[\exp(iuX_j)]= \{E[\exp(iuX_1)]\}^j,$$

because the X_k are independently and identically distributed. Then,

$$E[\exp(iuX_1)]=\int_{-\infty}^{\infty}\exp(iuX_1)dV=\int_{0}^{\infty}\exp(iuX_1)dV= \int_0^{\infty}\lambda\exp(iux-\lambda x)dx=\lambda/(\lambda-iu),$$

where the second step is possible by (2.6). But

$$\phi(u)=E[\exp(iuT_j)]=(1-iu/\lambda)^{-j},$$

is the characteristic function of the gamma distribution, thereby proving (2.7) by the inversion theorem.

In summary, the Todorovic and Woolhiser model consists of (2.1), (2.3), and (2.5), with $V(x)$ and $Pr(T_j \le x)$ given by (2.6) and (2.7).

Assumptions used by the model are:

1. the process Y_t is a first-order, two-state Markov chain,
2. the X_k are independently and identically distributed, and
3. the X_k and T_n are independent of s.

The second assumption means that the amounts of precipitation, on a series of wet days, are conditionally independent. The third assumption, physically speaking, means that a knowledge of s, the number of wet days in the n-day period, does not contribute any information about the daily amounts of precipitation X_k, or the total amount of precipitation T_n.

3.4 The Katz Model

Katz (1977a) generalized the recurrence relation approach to obtain recurrence equations for the maximum daily and total amount of precipitation in an n-day period.

Again, the sequence $\{Y_t;\ t=1,2,\ldots\}$ is assumed to be a first-order two-state Markov process. The distribution of the X_t is assumed to depend on Y_{t-1}, but the X_t are conditionally independent, given the Y_{t-1} process. The first part of the assumption means that, given a wet day, the amount of precipitation X_t is distributed according to $F_i(x)$, $i=Y_{t-1}$. The second part of the assumption means that knowledge of the amount of precipitation on a wet day does not contribute any knowledge about the amount of precipitation on other wet days.

3.4.1 Maximum Daily Precipitation

We define two conditional distributions for the maximum daily amount by:

$$G_n(x;i) = \Pr(M_n \le x \mid Y_0 = i),\quad i=0,1.$$

Conditioning on Y_0 gives

$$G_n(x) = (1-p)G_n(x;0) + pG_n(x;1). \tag{2.8}$$

Further conditioning, on Y_1, leads to

$$G_n(x;0) = p_{00}G_{n-1}(x;0) + p_{01}G_{n-1}(x;1)F_0(x), \tag{2.9}$$

and

$$G_n(x;1) = p_{10}G_{n-1}(x;0) + p_{11}G_{n-1}(x;1)F_1(x), \tag{2.10}$$

where $F_i(x) = \Pr(X_t \le x \mid Y_{t-1} = i)$, $i=0,1$. The initial conditions are simply: in zero days the probability that no

precipitation occurs is one, despite the occurrence or non-occurrence of precipitation on the previous day, i.e.,

$$G_0(x;0)=G_0(x;1)=1. \qquad (2.11)$$

The recurrence relations (2.9) and (2.10) with initial conditions (2.11) and (2.8) can be used to numerically calculate the distribution for M_n.

3.4.2 Total Precipitation Amount

Recall $T_n=\sum_{k=0}^{s} X_k$ is the total amount of precipitation in n-days and $H_n(x)=Pr(T_n \le x)$ is the distribution function for T_n. Letting

$$H_n(x;i)=Pr(T_n \le x \mid Y_0=i),\ i=0,1,$$

we have, upon conditioning on Y_0,

$$H_n(x)=(1-p)H_n(x;0) + pH_n(x;1). \qquad (2.12)$$

Conditioning on Y_1 gives

$$H_n(x;0)= p_{00}H_{n-1}(x;0) + p_{01}f_0{*}H_{n-1}(x;1) \qquad (2.13)$$

$$H_n(x;1)= p_{10}H_{n-1}(x;0) + p_{11}f_1{*}H_{n-1}(x;1) \qquad (2.14)$$

where * denotes the convolution

$$\int_0^x f_i(t)H_{n-1}(x-t;1)dt,\ i=0,1,$$

and $f_i=dF_i/dx$ is the density function for the daily precipitation amount. The initial condition is that the total amount of precipitation that can occur in zero days is zero, i.e.,

$$H_0(x;0)= H_0(x;1)= 1,\ x \ge 0. \qquad (2.15)$$

The convolutions appear in (2.13) and (2.14) because the probability of t amount of precipitation on the first

day is

$$p_{i1} f_i(t) dt,$$

the probability of x or less precipitation in the n days is

$$p_{i1} f_i(t) H_{n-1}(x-t;1) dt,$$

and since zero to x precipitation amount is possible on the first day, the probability of x or less precipitation in the n-day period is

$$p_{i1} \int_0^x f_i(t) H_{n-1}(x-t;1) dt = p_{i1} f_i * H_{n-1}(x;1).$$

The recurrence relations (2.13), (2.14) and initial conditions (2.15) can be used with (2.12) to calculate the cumulative distribution for T_n.

3.4.3 Application

The recurrence relation approach, proposed by Katz (1974, 1977a), can be used to calculate distributions for s, M_n, and T_n. In summary, the assumptions used in this approach are:

1. Y_t is a first-order, two-state Markov chain,
2. the distribution of X_t depends on Y_{t-1}, and
3. the X_t's are conditionally independent, given Y_{t-1}.

Unlike the Todorovic and Woolhiser model, the density function, f_i, need not be restricted to those having an analytical solution for the distribution of the sum of the stochastic variables.

The gamma distribution has often been selected to approximate the distribution of precipitation amount

occurring during a year, month, or day. (Skees and Shenton, 1971; Schickedanz and Krause, 1970). Since the gamma distribution is a common choice, and was selected by Katz, the gamma distribution was chosen in this study to approximate the distribution of daily precipitation amount in the Katz model.

CHAPTER 4.

Estimation of Parameters

4.1 General

Stochastic models generally require input in the form of a set of parameters estimated from the development data. The characteristics of the modeled process are imparted to the model through the input parameters. The set of parameters required for the models used here is

$$\Omega=(p,\ p_{10},\ p_{00},\eta,\lambda).$$

The parameter space consists of initial and transition probabilities for the occurrence of precipitation, and the shape and scale parameters for the distribution of daily precipitation amount. In general the parameters may be nonstationary, exhibiting temporal variations within a year, over a number of years, or both. The parameters for any given day of the year are assumed stationary over the years, but day-to-day nonstationarity of the parameters is recognized and allowed for.

4.2 The Markov Chain Parameters

Estimation of the initial and transition probabilities requires totalling the number of wet-dry day sequences occurring in the development data. The number of wet-dry day sequences is denoted by $n_{ij\ldots lm}(t)$, the observed daily frequency of the k transitions $i\to j\to\ldots\to l\to m$ ending in state m on day t, $1\le t\le 365$. The k+1 indices i, j,..., l, m denote

the states (0, 1) of the sequence $\{Y_t\}$ on the k+1 days ending on day t. The frequencies $n_{ij...lm}(t)$ of the development data were obtained using program COUNT which is listed in Appendix B. The N years of development data provide N independent observations for the $n_{ij...lm}(t)$. Frequencies were obtained for k=0,1,2,3,4. For k=0 the $n_i(t)$ are the unconditional number of wet (i=1) or dry (i=0) days occurring in the sample. The transition frequencies $n_{ij}(t)$, obtained with k=1, are the number of i to j transitions between days t-1 and t. The higher order frequencies, k=2,3,4, were obtained to test for the correct Markov chain order.

Leap years pose a problem for the sequence tabulation. A large sampling fluctuation can be expected for February twenty-ninth because of the few observations available. So the twenty-ninth of February was utilized in determining the sequence totals for March first to fourth, but sequences ending on February twenty-ninth were not tabulated. Yevyevich (1972) noted that this results in a one-quarter day shift in the period of each of the first three years and a three-quarter day shift in the fourth year following a leap year. Most results are not thought to be affected significantly by this problem (Yevyevich, 1972).

Gabriel and Neumann (1962), Hopkins and Robillard (1964), and Feyerherm and Bark (1964, 1965) suggested that the initial and transition probabilities should be allowed to vary during the year. Since daily variation of the

probabilities was easily incorporated into the recurrence relation method, the probabilities were calculated on a daily basis. Unfortunately, Gabriel's (1959) derivation for the probability of the number of wet days in an n-day period requires a homogeneous chain, i.e., constant transition probabilities. Since a month was thought to be a reasonable time period for which a calculated distribution would be useful, the transition probabilities were assumed constant within months when used in the TW model. Such an assumption may bias the results (Feyerherm and Bark, 1965), but the assumption was necessary and its validity will be examined in Chapter 5.

Daily and monthly initial and transition probabilities were estimated using the wet-dry-day sequence totals. The maximum likelihood estimates for the $p_{ij\ldots lm}$, the probability of the k transitions $i \rightarrow j \rightarrow \ldots \rightarrow l \rightarrow m$, were used. The daily transition probabilities $p_{ij\ldots lm}(t)$ are given by

$$p_{ij..lm}(t)=n_{ij..lm}(t)/\sum_{m=1}^{2} n_{ij\ldots lm}(t). \qquad (4.1)$$

Monthly transition probabilities were calculated using

$$p_{ij\ldots lm}=\sum_{t} n_{ij\ldots lm}(t)/\sum_{t}\sum_{m=1}^{2} n_{ij\ldots lm}(t) \qquad (4.2)$$

where the summation over t was carried out over the selected month.

The raw daily transition probabilities estimated by (4.1) can be used in the Katz model. But Feyerherm and Bark (1965) suggested that improved estimates of the

probabilities can be obtained by representing them by a Fourier series with a fundamental period of one year. Yevyevich (1972) justified such a fundamental period on the basis of astronomical cycles.

The 365 raw estimates for the three independent initial and transition probabilities ($1-p$, p_{00}, p_{10}) were used to estimate the coefficients for three Fourier series of the form

$$Pr(t)=A + \sum_{h=1}^{M} A_h \cos(2\pi ht/365)+B_h \sin(2\pi ht/365)$$

by the standard method of least squares (subroutine FOUR).

Yevyevich (1972) and Feyerherm and Bark (1965) discussed a number of statistical procedures for determining which of the possible 182 harmonics should be retained in the Fourier series. The procedures are somewhat complicated and require a number of assumptions about the residuals which are difficult to check and may be violated (Feyerherm and Bark, 1965; Yevyevich, 1972).

A simple graphical approach recommended by Yevyevich (1972) was used to select the number of harmonics necessary for the Fourier series. The method is to first plot a relative cumulative periodogram. The plot consists of

$$P_M=[\sum_{h=1}^{M} (A_h^2+B_h^2)/2]/\sigma^2$$

versus the harmonic M, M=1,2,...,182. The variance explained by each Fourier component h is $(A_h^2+B_h^2)/2$ and σ^2 is the total variance of the series of raw transition or

initial probability estimates. The Fourier series amplitudes are not summed in order of decreasing magnitude so a large component may be included after a small one.

Selection of the maximum number of harmonics to include in each series is based on Yevyevich's (1972) observation that the relative cumulative periodogram will consist of two parts: a fast-rising portion representing the periodicities in the data, and a slowly-rising part due to sampling variation. The two parts are approximated by smooth curves that intersect at a point specifying, in general, a non-integral critical harmonic, M_c. The procedure is then to accept all harmonics smaller than M_c. Yevyevich (1972) has found that daily series are nearly always periodic with a critical harmonic in the range of one to twelve; therefore, the coefficients of harmonics one to twenty-one only were calculated for the cumulative periodograms.

Using Fourier series to represent the nonstationary parameter space $(p(t), p_{10}(t), p_{00}(t))$ reduced the number of parameters required from 3x365 to $2M_1+2M_2+2M_3+3$ where M_1, M_2, M_3 are the number of harmonics selected for the Fourier series representations of $p(t)$, $p_{10}(t)$, and $p_{00}(t)$. A second benefit was the reduced variance of the Fourier series estimates for the initial and transition probabilities. According to Feyerherm and Bark (1965) the variance of the Fourier series estimate is reduced by a factor of $(2M_i+1)/365$ from that of the raw estimates, a significant amount for typical values of M_i, i=1, 2, 3.

Woolhiser and Pegram (1979) pointed out two drawbacks to using the method of least squares for estimation of Fourier coefficients. First, a varying sample size or varying properties of the distribution being fitted can result in unequal variances of the raw estimates. The method of least squares incorrectly gives each raw estimate equal weight. Second, there is no statistically sound procedure to test the significance of individual harmonics. Richardson (1977) indicated that the inclusion of too many harmonics perpetuates sampling error in the parameters while selection of too few harmonics results in an inaccurate description of the periodic nature of the precipitation process.

The cumulative periodogram method of harmonic selection was chosen over the alternatives suggested by Yevyevich (1972), Feyerherm and Bark (1964), and Woolhiser and Pegram (1979) because of its simplicity, and its intuitive appeal. Although there was a risk of selecting an incorrect number of harmonics, particularly when the transition from the quickly rising periodic part of the periodogram to the slowly rising sampling fluctuation portion occurred smoothly, the other procedures offered no guarantees of selecting the correct number of harmonics. To implement the maximum likelihood procedure suggested by Woolhiser and Pegram, in order to account for the unequal variance of the raw estimates and to select the Fourier series harmonics, was thought to be too time consuming for this work.

4.3 The Gamma Distribution Parameters

Selection of the gamma distribution

$$F_i(x) = \int_0^x \frac{\lambda_i^{\eta_i}}{\Gamma(\eta_i)} \tau^{\eta_i - 1} \exp(-\lambda_i \tau) d\tau, \quad i = 0,1 \qquad (4.3)$$

to represent the cumulative distribution for daily precipitation amount necessitated estimation of the shape η_i and scale λ_i parameters. Recall that the $F_i(x)$ in Katz's model were selected for day t such that $i = Y_{t-1}$.

Yevyevich (1972) claimed that the η_i and λ_i are nonstationary. Ison et. al. (1971) and Woolhiser et. al. (1973) found that the scale parameter λ_i had a seasonal variation and they accounted for the variation with Fourier series. The use of Fourier series for the gamma distribution parameters was rejected for the present study. To obtain good shape and scale parameter estimates a reasonably large sample of precipitation amounts was required. The number of wet days in the approximately fifty years of development data available for estimation of the parameters was expected to be too small, if short time periods of a day or week were used to obtain raw parameter estimates, particularly for dry seasons or stations with few wet days. Since the longest n-day period for which distributions were to be calculated was one month the shape and scale parameters were assumed constant within months. A month was thought to be sufficiently long to obtain large enough samples that

reliable parameter estimates could be obtained, but this was not checked. The month-to-month variation in the estimates accounted for the seasonal variation, yet was simpler than the Fourier series approach with its attendant problems.

The daily precipitation amounts in each month were abstracted from the development data by a computer routine ABSTR which sorted the precipitation amounts according to the occurrence (i=1) or nonoccurrence (i=0) of precipitation on the previous day. ABSTR also calculated the statistics required for a number of maximum likelihood techniques that were used for estimation of the shape and scale parameters.

Given the precipitation amounts, assumed to be a set of independent observations $\{X_{ij};\quad j=1,\ldots,N\}$ distributed according to (4.3), the maximum likelihood estimates for the shape parameter η_i and scale parameter λ_i were obtained by solving

$$\log \hat{\eta}_i - \psi(\hat{\eta}_i) = \log (\overline{X}_i/\overline{X}_{iG}) \tag{4.4}$$

$$\hat{\eta}_i/\hat{\lambda}_i = \overline{X}_i \tag{4.5}$$

where $\overline{X}_i$, $\overline{X}_{iG}$ are the sample arithmetic and geometric means, and $\psi(y) = d\log\Gamma(y)/dy$ is the psi or digamma function.

The set (4.4), (4.5) was not solved explicitly because of the complexity of the digamma function. Instead iterative numerical techniques have been used to solve the equations (Ison et. al., 1971). Mielke (1976) provided an iterative procedure for evaluating (4.4) exactly. Two variations of the procedure were given to accomodate a maximum likelihood ratio test on the scale parameters of two gamma

distributions with a common shape parameter (Schickedanz and Krause, 1970). The test, discussed in Chapter 5, was used to determine if the assumed difference, by the Katz model, between the distributions $F_0(x)$ and $F_1(x)$ was statistically significant.

Digressing to the test for a moment, because of its relevance to Mielke's procedure, let

$$\{X_{0j}; j=1,2,\ldots,N_0\} \text{ and } \{X_{1j}; j=1,2,\ldots,N_1\}$$

represent sets of N_0 and N_1 observations from gamma-distributed populations 0 and 1. The shape and scale parameters for population 0 and 1 are then denoted η_0, λ_0 and η_1, λ_1. The test constructed by Schickedanz and Krause (1970) tests the null hypothesis

$$H_o : \lambda_o = \lambda_1 = \lambda, \quad \eta_o = \eta_1 = \eta$$

against the alternate hypothesis

$$H_a : \lambda_o, \lambda_1, \eta_o = \eta_1 = \eta,$$

i.e., $\lambda_0 \neq \lambda_1$ in general. The likelihood ratio statistic, $\omega = L'/L$, is given by the ratio of the maximum likelihood under H_o to that of its largest possible value, under H_a (Kendall and Stuart, 1967). Schickedanz and Krause stated that $-2\log\omega$ is approximately distributed as a chi-square variate with one degree of freedom.

Mielke's procedure was used to calculate shape and scale parameters under H_o and H_a. Mielke claimed that the procedure results from (4.4), (4.5), and the approximation to the digamma function

$$\psi(\eta) \simeq -C + (\eta-1)\sum_{j=1}^{NS} [1/j(j+\eta-1)] + \log[(NS+\eta-\tfrac{1}{2})/(NS+\tfrac{1}{2})],$$

but gave no details of the derivation. The constant C is Euler's constant and NS is a selected integer (25) determining the accuracy of the digamma approximation. The shape and scale parameters were calculated under H_o, given an initial value of η, by

$$\eta_K = 1 + \frac{\log\left[\frac{\eta_{K-1}(NS + \frac{1}{2})}{NS + \eta_{K-1} - \frac{1}{2}}\right] + C - A}{\sum_{j=1}^{NS}\left[1\big/(j + \eta_{K-1} - 1)j\right]} \tag{4.6}$$

and

$$\lambda_K = \eta_K/\bar{X}$$

where

$$A = \log \bar{X} - \left(\sum_{j=1}^{N_0} \log x_{0j} + \sum_{j=1}^{N_1} \log x_{1j}\right)/N, \tag{4.7}$$

$$\bar{X} = \left(\sum_{j=1}^{N_0} x_{0j} + \sum_{j=1}^{N_1} x_{1j}\right)/N, \tag{4.8}$$

and $N=N_0+N_1$. Similarly, under H_a, η_k was given by (4.6),

$$\lambda_{0K} = \eta_K/\bar{X}_0, \quad \text{and} \quad \lambda_{1K} = \eta_K/\bar{X}_1,$$

where

$$A = (N_0 \log \bar{X}_0 + N_1 \log \bar{X}_1 - \sum_{j=1}^{N_0} \log x_{0j} - \sum_{j=1}^{N_1} \log x_{1j})/N$$

and

$$\bar{X}_0 = \left(\sum_{j=1}^{N_0} x_{0j}\right)/N_0, \quad \bar{X}_1 = \left(\sum_{j=1}^{N_1} x_{1j}\right)/N_1 . \tag{4.9}$$

Parameter calculation and application of the test were done with program GAM2 (Wong, 1980).

A number of other approximations to (4.4) were used to obtain a maximum likelihood estimate for the shape parameter. In some cases the approximations were applied to the data sorted by ABSTR and different parameter estimates were obtained for the distributions of X_0 and X_1. In other cases

the X_0, X_1 data were pooled to give a single data set for which parameter estimates were obtained. The procedure used to solve (4.4) was selected on the basis of Schickedanz and Krause's test, and will be given with the estimates for each case in the following section.

The first approximation used was Thom's (1958) solution of (4.4) that was based on the truncation of a series expansion for $\Psi(\eta)$. The shape parameter was given by

$$\hat{\eta} = (1+\sqrt{1+4A/3})/4A - \Delta\hat{\eta}, \tag{4.10}$$

where $\Delta\hat{\eta}$ is a correction for the series truncation and A, as given by (4.7), was used. The first term of (4.10) was evaluated and then the tabulated $\Delta\eta$ (Haan, 1977) was applied to obtain a final estimate. The scale parameter was obtained from (4.5) with $\bar{X}$ given by (4.8).

Greenwood and Durand's (1960) fraction approximation,

$$\hat{\eta} = (0.5000876+0.1648852A-0.0544274A^2)/A \tag{4.11}$$

for $0 \leq A \leq 0.5772$, and

$$\hat{\eta} = \frac{8.898919+9.05995A+0.9775373A^2}{A(17.79728+11.968477A+A^2)} \tag{4.12}$$

for $0.5772 \leq A \leq 17.0$, was also used. A was given by (4.7), and (4.5) was used to estimate the scale parameter. Greenwood and Durand claimed that the maximum error in (4.11) is 0.0088%, in (4.12) 0.0054%.

Haan (1977) claimed that the maximum likelihood estimates given by (4.10), (4.11), and (4.12) have a slight asymptotic bias and that the bias may be appreciable when only small samples are available. Estimates for the bias in the shape parameter $\hat{\eta}$ were given by Bowman and Shenton

(1968). According to Haan (1977), they suggested a simple approximation for the bias

$$E(\hat{\eta}-\eta)=3\hat{\eta}/N$$

which was rewritten

$$E(\eta)=(N-3)\hat{\eta}/N. \qquad (4.13)$$

Eq. (4.13) was used to correct for the bias in $\hat{\eta}$ when calculated by (4.10), (4.11), or (4.12).

The final approach used to estimate parameters for the gamma distribution attempted to account for trace rainfall. Traces represent a part of the precipitation process, and possibly the inclusion of all data available, i.e., traces, may provide better parameter estimates. The applicability of such an approach is questionable. A wet day has been defined as one on which a measurable amount of precipitation fell, i.e., more than a trace. But since the purpose of this work was to examine two models, rather than develop working models, the parameter estimates proposed by Das (1955) were used. The estimation procedure, as summarized by Skees and Shenton (1971), is followed here.

The distribution is the same as (4.3) except that it is truncated at $x=\epsilon$ where $\epsilon>0$ is small. The number of observations falling in the interval $(0,\epsilon)$ is T where ϵ was 0.13mm (0.005in prior to 1976). The total number of observations is N=T+m where m is the number of observations with precipitation amounts greater than ϵ. A Thom-type approximation was constructed

$$\hat{\eta}=1-2\theta+\sqrt{((1-2\theta)^2+4y/3)}\,/4y$$

where $y=\log\bar{X}-\overline{\log X}-\theta\log\epsilon$, $\overline{\log X}=(\sum_{j=1}^{m}\log x_j)/N$, $\theta=T/N$, and

$$\bar{X}=(\sum_{j=1}^{m}x_j)/N, \tag{4.14}$$

The scale parameter λ was calculated using (4.5) with $\bar{X}$ given by (4.14).

The exponential distribution (2.6), used to represent the distribution of daily precipitation amounts by the TW model, is simply the two-parameter gamma distribution with a shape parameter η of exactly one. Estimates for the scale parameter λ were obtained with (4.5) and (4.8).

4.4 The Estimates

4.4.1 Markov Chain Parameters

The monthly transition probabilities for use by the TW model, as given by (4.2), are summarized in Table 1. Two estimates for the initial probabilities, p, are given for each case. The first, p, was calculated using (4.2). The second, p′, is a Fourier series estimate for the day previous to the first day of the case month, e.g., December thirty-first for the Edmonton (January) case.

Subroutine FOUR, listed in Appendix B, was used with COUNT to calculate the amplitudes of the Fourier series harmonics, to calculate the relative cumulative variance of the harmonics, and to plot the cumulative periodograms for 1-p, p_{00}, and p_{10}. The cumulative periodograms for Beaverlodge, Figures 2 and 3, show that four, zero, and four harmonics of the Fourier series for 1-p, p_{10}, and p_{00} respectively should be included. For Edmonton, Figures 4

and 5 show that five, two, and two harmonics should explain the periodic nature of the probabilities 1-p, p_{10}, and p_{00}. Three, four, and two Fourier series harmonics were included in the series estimates for 1-p, p_{10}, and p_{00} at Medicine Hat, on the basis of Figures 6 and 7. The amplitudes of the harmonics selected are given in Table 2.

The cumulative periodograms for Edmonton, Figures 4 and 5, exhibit a sharp transition from the quickly rising portion of the curve to the slowly rising section. Choosing the number of harmonics for inclusion in the series estimates for the transition probabilities was straightforward. The decision to include the third to fifth harmonics of the series for 1-p was more difficult, but was justified by Figure 4. Inclusion of the third to fifth harmonics explained an additional five percent of the raw estimate's variance.

The smooth transition from the periodic to residual variance portions of the cumulative periodograms for Beaverlodge and Medicine Hat presented a problem. A straight line or smooth curve was fitted so that it seemed to represent the periodic portion of the periodogram. Curve fitting was done by eye and was quite subjective.

The Fourier series estimates for the daily initial and transition probabilities were plotted with their raw daily estimates for each case. Figures 8 to 16 show that the Fourier series provide estimates in reasonable agreement with the observed raw probabilities, as they should since

the raw estimates were used to calculate the Fourier series coefficients. But Figures 8 and 14, and 9, 12, and 15, suggest that the Fourier series overestimate the probabilities 1-p and p_{00} for their respective cases in June and early July (approximately days 150 to 190). The overestimate is simply the result of the least squares fit, the estimate would not be significantly reduced by inclusion of an additional harmonic. In particular, the addition of another harmonic for p_{00} at Edmonton (Figure 12) could not be justified in the light of Figure 5. The effect of the overestimation of p_{00} will become evident in Chapter 6.

4.4.2 Gamma Distribution Parameters

The shape and scale parameters estimated using Mielke's procedure are summarized in Table 3. The table also includes the significance levels achieved by Schickedanz and Krause's likelihood ratio test under the null hypothesis of equal scale parameters for $F_0(x)$, $F_1(x)$.

The null hypothesis was not rejected for the Edmonton and Medicine Hat cases at the 0.10 significance level--the probability, given in Table 3 by α, of a random chi-square variate with one degree of freedom exceeding the calculated chi-square value was greater than 0.10. But the null hypothesis was rejected at the 0.05 significance level for both Beaverlodge cases. The test indicated that there was no difference between $F_0(x)$ and $F_1(x)$ for the Edmonton and Medicine Hat cases, assuming the shape parameters were the

same. For the Beaverlodge cases the test showed that $F_0(x)$ and $F_1(x)$ were significantly different.

The appropriate Mielke scale parameters were used for $F_0(x)$ and $F_1(x)$ in the Katz model for the Beaverlodge cases. Das parameter estimates for $F_0(x)$ and $F_1(x)$ at Beaverlodge are given in Table 4. These estimates were obtained from data sorted according to the wet-dry state of the day previous to the observed amounts.

Mielke's estimates were not used for Edmonton and Medicine Hat. Instead, the parameters calculated from the approximate methods given, and pooled data (the data for day t, originally sorted by ABSTR according to the wet-dry state of day t-1, were pooled to give a single set of data for each case) were used for both $F_0(x)$ and $F_1(x)$ in the Katz model for the four cases. The estimates of the shape and scale parameters for Edmonton and Medicine Hat, calculated using the methods of Thom (1958), Greenwood and Durand (1960), and Das (1955), are summarized in Table 5. Table 6 contains the scale parameters estimated for the exponential distribution. The pooled data for each case, including the Beaverlodge cases, were used to calculate these scale estimates.

Table 5 shows that the methods of Thom, and Greenwood and Durand provided essentially the same shape parameters, and so the scale parameters were taken to be the same. Comparison of Tables 3 and 5, for the Medicine Hat and Edmonton entries, shows that Mielke's iterative procedure

provided estimates that confirm the Thom and Greenwood-Durand (TGD) estimates for the shape parameter, prior to the correction for bias. This suggests that the Mielke parameter estimates used for the Beaverlodge case are biased; this result was expected because the different techniques all solve (4.4). The parameters given in Table 3 for the Beaverlodge case were not corrected for bias before use.

In summary, the parameter estimates used in the Katz model for the Beaverlodge case included the Mielke estimates in Table 3 and the Das estimates in Table 4. For the Edmonton and Medicine Hat cases, the TGD and Das estimates in Table 5 were used for the Katz model. The TW model used the parameters in Table 6 for all cases.

The theoretical and observed distributions for $F_0(x)$ and $F_1(x)$ are shown in Figures 17 and 18, and 19 and 20, for May and July at Beaverlodge. The distributions of the pooled data for May and July at Beaverlodge are shown with the exponential distribution in Figures 21 and 22. The observed and theoretical distributions for Edmonton and Medicine Hat are shown in Figures 23 to 26.

4.5 Goodness of Fit

The goodness of fit of the gamma distribution to the observed distribution of daily precipitation amount was determined by a visual judgment and application of the Kolmogorov-Smirnov (K-S) test. A visual judgment of the fit was obtained by comparing plots of the observed and

theoretical distributions. The K-S test was used to test the null hypothesis that the observed and theoretical distributions were the same. The statistic

$$D_N = \max|F(x) - O(x)|$$

was calculated and compared to critical values given by Crutcher (1975) for use when parameters are estimated from the observed data. F(x) was the theoretical gamma distribution and O(x) was the observed distribution.

The fit of the gamma and exponential distributions to the observed distributions, for the daily amount of precipitation, was generally reasonable, but not good. In only three instances, the Beaverlodge May case for $F_0(x)$ and $F_1(x)$ and the July case for $F_0(x)$, did the theoretical gamma distribution with Das parameters closely follow the observed distribution for low precipitation amounts, where the observed distributions exhibited steep slopes. In all cases the theoretical distributions over-estimated the observed distribution for the larger precipitation amounts observed. In all cases the distributions using the TGD or Mielke parameters under-estimated the observed distributions for the smaller precipitation amounts observed and over-estimated the observed distribution for the larger amounts. The exponential distribution did the same. For the Edmonton and Medicine Hat cases the distributions using Das parameter estimates over-estimated the observed distributions for the range of precipitation amounts observed.

The Das parameter estimates for the four Beaverlodge

cases provided the best fits for the smaller precipitation amounts observed. The gamma distribution closely followed the observed in the zero to six or eight millimetre range, but then began to deviate, more so for the July case than for the May case.

Figures 24 and 26 show that the exponential and gamma distributions using the TGD estimates were nearly the same, for June at both Edmonton and Medicine Hat; for January at Edmonton and March at Medicine Hat they were identical (Figures 23 and 25).

On the basis of Figures 17 to 26 the gamma distributions exhibiting the best fits were those using the Das parameter estimates for the May case at Beaverlodge and the June case at Edmonton. Consequently, the derived distributions for the maximum daily and total amount of precipitation for those cases were expected to show better agreement with the observed distributions, because of better input about the distribution of daily amounts.

The Kolmogorov-Smirnov statistic, D_N, is given in Tables 3, 4, 5, and 6 for each of the parameter estimates used in the gamma and exponential distributions. For each case's parameter sets, with two exceptions, the null hypothesis that the theoretical gamma and observed distributions were the same was rejected at the 0.05 level of significance. Because Crutcher (1975) did not provide critical values for D_N when non-integral values of η are estimated, the K-S test was first applied using the non-parametric

critical value of $1.36/\sqrt{N}$, where N is the number of observations.

The null hypothesis was rejected for each set of parameters for the Edmonton and Medicine Hat cases. Since the K-S test is conservative with respect to Type I errors when parameters are estimated from the data, the true significance level of the test was less than 0.05 (Crutcher, 1975). In other words, the null hypothesis was rejected with considerable confidence for the Edmonton and Medicine Hat cases.

For all cases, the exponential distribution was significantly (0.05) different than the observed distribution of daily precipitation amount.

Similarly, the null hypothesis was rejected for both Beaverlodge cases when the Mielke parameter estimates were used, and when the Das parameter estimates were used for the distribution of precipitation amount on days following a dry day in July.

The gamma distribution was not found to be significantly different from the distribution of observed daily precipitation amount with the previous day wet during July, when a non-parametric critical value at the 0.05 level of significance was used. But, using the parametric critical value (η estimated and equal to 1) supplied by Crutcher, $1.05/\sqrt{N}$, the null hypothesis was rejected. Since the shape parameter was not equal to one, the former result was accepted, because the null hypothesis was just rejected when

the parametric critical value was used.

The only parameter estimates calculated for the gamma distribution, such that the observed and theoretical distributions were clearly the same on the basis of the K-S test, were the Das estimates for May at Beaverlodge. The null hypothesis was not rejected for either distribution, previous day wet or previous day dry, using Crutcher's (1975) critical value of $1.05/\sqrt{N}$.

If an operational model had been the objective of this work, an attempt to adjust the parameters to give a best possible fit in all cases would have been made. But to determine, if possible, the influence that the distribution of daily precipitation amount had on the distributions for the maximum daily and total amount of precipitation the parameter estimates given in Tables 3 to 6 were used.

Use of the chi-square goodness of fit test was discouraged by a possible measurement bias in the data. The bias was to precipitation amounts that were multiples of one-tenth of an inch. No statistical test was used to determine if the number of observations of 2.5mm (0.10in) or 5.1mm (0.20in) of precipitation was excessive. But the numbers are suggestive of a bias, particularly in the development data for the Edmonton January and Medicine Hat March and June cases. Table 7 gives the number of times 2.5mm, 5.1mm and the two amounts adjacent to them were recorded in the development data for each case, and for March at Beaverlodge.

The bias in the Edmonton January and Medicine Hat March development data was quite possibly the result of observers measuring snowfall to the nearest inch and dividing by ten to obtain a water equivalent. The record for March at Beaverlodge had a striking example of such a bias. The apparent decrease in the bias of the Beaverlodge record in warmer months was probably because improperly trained observers are more able, or more willing, to record non-round numbers obtained with a rain gauge and graduate cylinder than with a snow ruler. But the existence of a bias toward 0.10in and 0.20in in warmer (rain) months is still evident in Table 7; only one of the six summer-month amount combinations (Beaverlodge July, at 5.1mm) did not have a maximum number of observations for a multiple of 0.10in. Other maxima in the observed frequency of precipitation amounts were found for 0.30in, 0.40in, and 0.50in, but the maxima were not as striking because of the fewer occurrences of the larger precipitation amounts. The probability of obtaining the arrangement in Table 7, given fourteen independent sets of three numbers, each set arranged in an equiprobable fashion is $(1/3!)^{13}(4/3!)$.

The bias, in some cases, resulted in unrealistically large contributions to the chi-squared statistic when a test of fit was attempted; the effect was to reject the null hypothesis that the distributions were the same. The bias, and the difficulty in selection of class intervals, were the reasons for not using the chi-square goodness of fit test.

CHAPTER 5.

The Assumptions

5.1 General

A critical examination of the assumptions required by the models and inherent in modeling a climate record, is necessary. Without such an examination the applicability of the model chosen cannot be determined; misleading or erroneous results may be incorrectly accepted. The methods given in this chapter were used to examine the assumptions required for the theoretical development of the models, for parameter estimation, and for climatic record modeling, in an attempt to detect breakdowns in the assumptions that may lead to improper results.

5.2 Stationarity

Both models assume, to some extent, that the precipitation time series $\{Y_t, X_t\}$ is stationary. The TW model requires stationary transition probabilities and identically distributed X_t, i.e., a stationary distribution. The Katz model requires the X_t process to be stationary within months. The procedure used to determine the correct Markov chain order needs a stationary $\{Y_t\}$ process, both within months and over the years because data observed in successive years were used. The assumption that the series distributions are constant over the years is also inherent in expecting a model using parameters from a single realization

of the precipitation time series, $\{Y_t, X_t\}$, to represent future realizations of the process. Accordingly, an attempt was made to ascertain whether or not the precipitation process was stationary.

To facilitate the study of hydrologic time series Yevyevich (1972) has identified two basic components to series structure. The first is deterministic; the second stochastic. The deterministic component may take the form of jumps, cycles, or long term trends. Trends or jumps may appear in the deterministic component because of inconsistency (systematic errors) or nonhomogeneity (changes in nature because of man, or natural causes) of the data.

Yevyevich has identified a periodicity with a fundamental of one year to be an important natural deterministic component that is nearly always present in hydrologic data. Feyerherm and Bark (1965) used Fourier series with a fundamental period of one year to model the changes in transition probabilities observed within a year. Their work motivated the use of Fourier series to account for daily changes in the transition probabilities in the Katz model used in this study. But daily changes in the transition probabilities were not permitted by the TW model. Figures 9, 12, 13, 15, and 16 show that the transition probabilities did vary within months; the test given here was used to determine if the variations were statistically significant.

Woolhiser et. al. (1973) and Kaavas et. al. (1977) have used the test to show that transition probabilities are

stationary during a week in eastern Colorado and at Ankara, Turkey. The time period required for the TW model was one month.

The test, constructed by Anderson and Goodman (1957), tested the null hypothesis that the transition probabilities were constant

$$H_0: p_{ij}(t) = p_{ij}$$

against the alternate hypothesis that they were nonstationary. The test used the maximum likelihood estimates for the transition probabilities given by (4.1) and (4.2) when N realizations of the process, each of length T, were available, i.e., N years of data with T equal to the number of days in the month considered. The likelihood ratio

$$\omega = \prod_{t} \prod_{i,j} [\hat{p}_{ij} / \hat{p}_{ij}(t)]^{n_{ij}(t)}$$

was calculated and $-2\log\omega$ was compared with a chi-squared variate with 2(T-1) degrees of freedom.

The null hypothesis could not be rejected at the 0.10 significance level for the Edmonton cases, the Medicine Hat cases, or the May case for Beaverlodge. The null hypothesis was rejected at the 0.01 significance level for July at Beaverlodge.

The test results show that the transition probabilities, within the case months examined here, can be considered constant, with the exception of the transition probabilities in July at Beaverlodge, which are not constant. The latter result suggests that the Katz model should

approximate the distribution of the number of wet days, in an n-day period in July at Beaverlodge, better than the TW model.

The distribution of the daily precipitation amounts, $\{X_t\}$, was assumed to be stationary within a month for both models. Because different arithmetic and geometric mean daily precipitation amounts were obtained for different months it was concluded that the distribution of daily precipitation amounts varied during the year. The climatic normals in Tables 8, 9, and 10 also support such a conclusion. Although the number of wet days in the summer months is occasionally greater than for other months, the normal monthly precipitation total is two to four times as large, indicating more rain on a wet day. It seemed reasonable to expect the distribution for the daily amount of precipitation to vary continuously throughout each month of the year. Such variation violated the assumption that the X_t's were identically distributed during a month. Whether or not the variation in the distribution for X_t, over a month-long period was statistically significant was not determined.

In short time series, long-term trends and cycles (over a number of years) in the deterministic component are often the result of sampling fluctuations. Determining the statistical significance of the cycles or trends is difficult. Yevyevich (1972) suggested that a historical study of factors possibly influencing the time series should be carried out to substantiate the statistical detection of

trends or jumps. To this end, the station histories compiled by Lachapelle (1977) were used in Chapter 2 to identify a time when a major change at the observing site occurred. In order to determine the statistical significance of any inhomogeneity or inconsistency induced in the data by the change, the tests were applied across the date of the change whenever possible.

Simple statistical techniques were used in the attempt to detect nonhomogeneity and inconsistency in the data. To detect long-term variation in the precipitation occurrence process, $\{Y_t\}$, a two-sample t-test was used to test for differences in the initial probabilities of $\{Y_t\}$. The temporal structure of the daily amounts, $\{X_t\}$, was examined using a two-sample z-test and linear regression on ten-year mean wet-day precipitation amounts. Normal monthly precipitation totals were also examined.

The maximum likelihood estimate for the probability of a wet day, p, is simply the mean of the random variate Y_t. For a sample of sufficient size the probability p should be normally distributed, according to the Central Limit Theorem. Large samples were used to ensure the stability and normality of the estimates for the mean p, of the U-shape-distributed random variate Y_t.

The entire record of observations at each location was split into two samples. The Beaverlodge samples were composed of the records from 1914 to 1957 and 1958 to 1978. The Edmonton samples encompassed the years 1883 to 1937 and

1938 to 1978. The Medicine Hat samples ran from 1884 to 1931 and 1932 to 1978.

Under the null hypothesis of equal means the two-sample t-statistic

$$t=(p_1-p_2)/\sqrt{\{[(N_1+N_2)/N_1N_2][(N_1-1)s_1^2+(N_2-1)s_2^2]/(N_1+N_2-2)\}}$$

with N_1+N_2-2 degrees of freedom was then obtained for each day of the year. The null hypothesis was rejected for any of the 365 days if $|t|$ exceeded $t_{\alpha/2,N_1+N_2-2}$. A large number of rejections of the null hypothesis—in excess of $365\times\alpha$, where α was the chosen level of significance—was evidence for the rejection of the assumption of long term stationarity of the precipitation occurrence process at the location considered, provided the tests are independent. The author is uncertain about the validity of this latter assumption.

At Beaverlodge, Edmonton, and Medicine Hat, there were twenty-nine, forty, and thirty-five days respectively for which the null hypothesis was rejected. At the five percent significance level used, eighteen to nineteen chance rejections of the null hypothesis were expected, even if it was true. The excessive number of rejections of the null hypothesis indicated that the initial probabilities for the precipitation process $\{Y_t\}$ were nonstationary for approximately ten percent of the days of the year at each location. In particular, there was evidence that the initial probability of precipitation had decreased at

Edmonton and Medicine Hat. In excess of seventy percent of the calculated t-statistics were negative for those locations while sixty percent of the calculated t-statistics were negative for Beaverlodge.

Breaking the test results down by cases, there were 7, 2, 4, 1, 3, and 3 days for which the null hypothesis was rejected for the: Edmonton, January and June, Medicine Hat, March and June, and Beaverlodge, May and July, cases. One or two rejections each month were expected when the null hypothesis was true. Therefore the initial probability of precipitation for June days at Edmonton and Medicine Hat seemed to be the same in both of their respective samples. The initial probability of precipitation in each sample was significantly (0.05) different for about ten percent of the days in the other case months, with the exception of the Edmonton-January case. In that case the null hypothesis was rejected for seven of thirty-one days, and only one t-statistic calculated was positive for the month.

The two-sample t-test provided evidence that supports the suggestion that the precipitation occurrence process $\{Y_t\}$ was nonstationary for approximately twenty-three percent of the days in January at Edmonton. The small number of positive t-statistics indicated that the probability of precipitation in January at Edmonton decreased from 1883-1937 to the levels observed in the 1938-1978 period. The t-test also indicated that the $\{Y_t\}$ process was nonstationary for approximately ten percent of the days in the

Medicine Hat-March and Beaverlodge cases.

However, the evidence for nonstationarity of the $\{Y_t\}$ process over the time periods considered for these cases was not overwhelming. And there was no evidence of nonstationarity in $\{Y_t\}$ for June at Edmonton and Medicine Hat.

On the basis of these results, the models' abilities to produce distributions representative of the independent data should not be unduly affected for all cases but the Edmonton-January case. There are difficulties in interpreting the two-sample t-test in terms of homogeneity of the $\{Y_t\}$ time series, and for completeness an examination of the transition probabilities should have been done. However, the test was performed to aid in the interpretation of the models' performance, and the results should do so. The test was not applied to the transition probabilities because a further subdivision of the data would have resulted in samples too small for reliable testing.

The assumptions required for application of the two-sample t-test include the normality of the parent population of the means, independence of the observations, and equality of the standard deviations. Although the distribution of Y_t is radically different from normal, the Central Limit Theorem should ensure near normality of the distribution of the means. Despite persistence of the random variate Y_t, the observations should be independent. To expect the value of Y_t on day t in one year to be dependent on Y_t of the previous year is unrealistic. The assumption of equal variance

for the parent populations was not examined. The t-test used is robust (Kendall and Stuart, 1967) and the test results should be valid.

The distribution of the daily precipitation amount for any given month has been assumed constant throughout the period of record. An attempt to find evidence of trends or jumps because of inconsistency or inhomogeneity in the data was made.

First, the long-term mean monthly precipitation totals published by the AES, and listed in Tables 8, 9, and 10, were examined for irregularities. The published monthly precipitation totals were converted to metric values and then divided by the published mean number of wet days in the appropriate month to obtain a long-term mean precipitation amount for a wet day during the month. The three values available for each case month were then compared.

The variation of the mean wet-day amount within cases was generally less than 1.0mm. For March at Medicine Hat and June at Edmonton the ranges of the values were 1.3mm and 1.0mm. The small range of the mean wet-day amounts was a coarse indication that the $\{X_t\}$ process was stationary in the mean.

Second, a two-sample z-test under the null hypothesis that the mean monthly wet-day precipitation amounts were equal, for successive ten-year means, was applied to detect jumps in the data. The z statistic

$$z=(X_1-X_2)/\sqrt{(\sigma_1^2/N_1+\sigma_2^2/N_2)}$$

was calculated for each pair of ten-year mean wet-day precipitation amounts and the null hypothesis rejected when $|z|>z_{\alpha/2}$, where α was the 0.05 level of significance. Sample estimates s_1^2 and s_2^2 for the variances σ_1^2, σ_2^2 were used. The ten-year means were expected to be normally distributed, by the Central Limit Theorem. The two-sample z-test is robust and the results should be valid.

The null hypothesis was not rejected for any pair of ten-year mean wet-day amounts for the Beaverlodge cases. The null hypothesis was rejected for a number of sample pairs in each of the following three cases.

At Edmonton, the mean wet-day amount for January in the 1888 to 1897 period was significantly larger than the succeeding means, but there is no historical note of a change at the Edmonton site at that time, and the sample size for that period was quite small. There were only thirty-six wet days in that time period while for most ten-year means there were more than one-hundred observations. The January mean for the 1968 to 1977 period was significantly smaller than the other means, but again, there is no historical evidence to support this result.

The ten-year mean for the 1928 to 1937 period in June at Edmonton was significantly smaller than a number of the other means calculated. This did not result from a site change since the 1928 to 1937 mean was significantly different than means for periods both before and after 1937.

There was no evidence that the $\{X_t\}$ process at Medicine

Hat in June had changed since 1892. The 1922 to 1931 mean was significantly lower than five of the means for other periods. But means for periods before and after the site change were larger than the 1922 to 1931 mean, so the change in mean for that period was not because of site changes.

Third, simple linear regression was performed on the successive ten-year means of wet-day amount, in an attempt to determine if a linear trend was evident in the precipitation amounts. The coefficients a and b of the relation

$$X=a+bT \tag{5.1}$$

where X is the ten year mean wet-day amount and T is the year were determined by least squares. The coefficient b,

$$b=\sum_{i-1}^{N}(T_i-\bar{T})(X_i-\bar{X})/\sum_{i-1}^{N}(T_i-\bar{T})^2,$$

where N is the number of ten year means, was used in a t-test of the null hypothesis

$$H_o: \beta=0$$

against

$$H_a: \beta\neq 0$$

where β is the population value for the slope of the trend. Acceptance of the null hypothesis was considered evidence that no linear trend existed in the mean wet day amounts.

The statistic $t=b/S_b$ was calculated and the null hypothesis rejected if $|t|>t_{\alpha/2,N-2}$. The variance of the estimate for b was given by

$$S_b=\sqrt{s^2/\sum_{i-1}^{N}(T_i-\bar{T})^2},$$

where s was the standard error of the regression,

$$s=\sqrt{\sum_{i-1}^{N} (e_i)^2/N-2} \ ,$$

and the e_i were the residuals $X_i - \hat{X}_i$. The $\hat{X}_i$ were given by (5.1) and the X_i were the observed amounts at time T_i (Haan, 1977).

The null hypothesis that β was equal to zero was not rejected for the Beaverlodge cases, the Medicine Hat cases, or the Edmonton June case. January at Edmonton exhibited a significant (0.05 level) linear trend of decreasing ten-year mean wet-day precipitation totals with time. No attempt was made to determine the exact cause of the latter result.

Assumptions required by the t-test are that: the e_i's were normally distributed with mean zero, uncorrelated, and homoscedastic with variance σ^2 (estimated by s^2). The first two assumptions were examined and found to hold, but the few points available made them difficult to check conclusively.

The controversy about natural long-term periodicity (Rodriguez and Yevyevich, 1967) or change (Clark, 1979) in climate makes prudent interpretation of the results a necessity. However, it is reasonable to claim that in general the test results supported the assumption that the $\{X_t\}$ process was stationary. A notable exception was the significant downward trend in ten-year mean daily precipitation amount in January at Edmonton. The trend in this case was not removed.

The techniques used to detect nonstationarity in the

distribution of daily precipitation amounts were crude. However, the test results may be of value in understanding the two models' abilities to reproduce distributions for the maximum daily and total precipitation in a given period.

Long term periodicities were assumed to be nonexistent in the data and no attempt was made to detect such a periodicity. Lachapelle (1977) found some evidence for the existence of a 10.7 year periodicity in the averaged June, July, and August monthly precipitation totals for Edmonton. Whether or not this periodicity affected the models' results for the June at Edmonton case is not known.

5.3 Markov Chain Order

The first-order Markov chain, because of its simplicity, is generally favoured in the literature on the stochastic modeling of the rainfall process. Nevertheless, Chin (1977) used an information theoretic decision criterion to show that the order of a Markov chain model of the precipitation occurrence process, $\{Y_t\}$, cannot be assumed a priori. A most crucial assumption of the models is that the stochastic process $\{Y_t\}$ constitutes a first-order Markov chain. Consequently, a determination of the appropriate Markov-chain-model order for the cases selected was deemed necessary.

The classical Neyman-Pearson theory of hypothesis testing is inadequate for model order selection (Gates and Tong, 1976; Katz, 1979a). Chin (1977) noted the loss caused

by the decision is inadequately defined as the probability of an error in incorrectly accepting or rejecting a particular model. The significance levels must be subjectively selected and there is no requirement for a simple model. Schwartz (1978) pointed out that since the maximum likelihood principle generally selects the highest possible order it cannot be the proper formalization of the intuitive notion of selecting the right order.

Two new approaches were used for chain order selection. Akaike (1971) suggested the first approach, extending the maximum likelihood principle to obtain an information theoretic criterion. Schwartz (1978) proposed the second, an alternate criterion based on a Bayesian argument and Katz (1979b) established the validity of the criterion for Markov chains.

The criteria adopt a parsimonious approach to model order selection, balancing the requirement for a good fit against increased complexity of the model. Both criteria balance two opposing terms: a log likelihood ratio and a penalty term which depends on the degrees of freedom of the model. The likelihood ratio statistic is the same for both criteria; similar looking, but fundamentally different penalty functions are used. Fitting higher-order models to the observed data reduces the log likelihood ratio, implying a reduction in residual variance (Gates and Tong, 1976). But the reduced variance is at the expense of a more complex model, indicated by the increased penalty. The best model

is the one having the minimum criteria value.

Akaike's information theoretic criterion (AIC) is defined

$$AIC(K)=-2\log(\text{maximum likelihood}) + 2K,$$

where K is the number of independent parameters in the model. The criterion is a measure of the difference between the true structure and the model, in terms of Kullback-Liebler information (Akaike, 1971).

Kullback (1959) defined the log of the likelihood ratio,

$$\log[f_1(x)/f_2(x)],$$

as the information in an observation x for discrimination in favour of H_1 against H_2. H_i, i=1,2, is the hypothesis that X is from a population with density function $f_i(x)$. The mean information for discrimination in favour of H_1 against H_2 per observation of X under H_1 was defined to be

$$f_1(x)\log[f_1(x)/f_2(x)]dx \text{ (Kullback, 1959).}$$

That $\log[f_1(x)/f_2(x)]$ is the information in observation x for discrimination in favour of H_1 against H_2 can be understood by considering Bayes Theorem

$$Pr(H_i|x)= \frac{Pr(H_i)f_i(x)}{Pr(H_1)f_1(x)+Pr(H_2)f_2(x)}, i=1,2,$$

where $Pr(H_i)$ is the prior probability of H_i, and $Pr(H_i|x)$ is the posterior probability of H_i after observation x. Then

$$\log[f_1(x)/f_2(x)]=\log[Pr(H_1|x)/Pr(H_2|x)]-\log[Pr(H_1)/Pr(H_2)]$$

is a measure of the difference between the log of the odds in favour of H_1 after the observation x and the log of the odds in favour of H_1 before the observation (Kullback,

1959).

Akaike (1971) began with the result (Blackwell, 1953) that the necessary information for discrimination between two probability distribution functions with density functions $f_1(x)$ and $f_2(x)$ is contained in the likelihood ratio $f_1(x)/f_2(x)$. He then showed Kullback-Liebler's definition of information was appropriate and extended the maximum likelihood principle to obtain the AIC.

Tong (1975) proposed the loss function

$$R(K)={}_K I_Q - 2(S^Q - S^K)(S-1), \tag{5.2}$$

based on the AIC approach, for use in identifying the Markov chain order of a process. The highest-order model considered is Q, K is the model order being tested, S is the number of states, and ${}_K I_Q$ is the log likelihood ratio statistic. The model, among those possible, that minimizes the loss R(K) is selected.

The second criterion, termed the Schwartz Bayesian criterion (SBC) by Katz (1979a), is defined

$$SBC(K)={}_K I_Q - (S^Q - S^K)(S-1)\log N, \tag{5.3}$$

where N is the sample size. That model minimizing SBC(K) is selected. The fundamental difference between the penalty terms of the two criteria is the inclusion of the sample size in the SBC estimator.

Gates and Tong's (1976) development of the likelihood ratio statistic is now summarized, followed by comments on application of the criteria.

The probability, or likelihood, of obtaining the

observed sequence $Y=\{Y_1, Y_2, \ldots, Y_N\}$ is

$$Pr(Y)=Pr(Y_1)Pr(Y_2|Y_1)Pr(Y_3|Y_2,Y_1)\ldots Pr(Y_N|Y_{N-1}\ldots Y_1)$$

so

$$L=Pr(Y_1)\prod_{\nu=2}^{N} Pr(Y_\nu|Y_{\nu-1}\ldots Y_1).$$

For a chain of at most order K,

$$L=Pr(Y_1)Pr(Y_2|Y_1)\ldots Pr(Y_K|Y_{K-1}\ldots Y_1)\prod_{\nu=1}^{N-K}Pr(Y_{K+\nu}|Y_{K+\nu-1}\ldots Y_\nu).$$

The last term dominates for large N. Then

$$L \approx \prod_{ij\ldots lm} p_{ij\ldots lm},$$

where the first K terms are ignored and the transition probability

$$Pr(Y_{K+\nu}|Y_{K+\nu-1}\ldots Y_\nu)$$

of a K chain is again denoted $p_{ij\ldots lm}$.

The likelihood ratio statistic used in the criteria is an asymptotic version of the likelihood ratio test statistic for composite hypotheses (Gates and Tong, 1976). The appropriate null hypothesis, that the chain is of order K, is

$$H_K: p_{ij\ldots lm}=p_{ij\ldots lm}.$$

The alternate hypothesis, that the chain is of order K-1, is

$$H_{K-1}: p_{ij\ldots lm}=p_{j\ldots lm}.$$

Using the maximum likelihood estimates for the transition probabilities given by (4.2),

$$p_{ij\ldots lm}=n_{ij\ldots lm}/n_{ij\ldots l},$$

where $n_{ij\ldots l} = \sum_{m=1}^{2} n_{ij\ldots lm}$, and denoting estimates under H_{K-1}

by a prime,

$$p'_{ij\ldots lm} = p_{j\ldots lm} ,$$

the likelihood ratio test for testing H_{K-1} against H_K takes the form

$$\phi_{K-1,K} = L(p'_{ij\ldots lm})/L(p_{ij\ldots lm}) .$$

For normally distributed $n_{ij..lm}$, ${}_K I_{K-1} = -2\log\phi_{K-1,K}$ is asymptotically a chi-square variate with $S^{K-1}(S-1)^2$ degrees of freedom under the null hypothesis (Hoel, 1954). The $n_{ij\ldots lm}$ are asymptotically normally distributed if the chain is ergodic (Bartlett, 1951).

The test is applied by calculating and then comparing

$$-2\log\phi_{K-1,K} = 2 \sum_{ij\ldots \ell m} n_{ij\ldots \ell m}\left(\log\frac{n_{ij\ldots \ell m}}{n_{ij\ldots \ell}} - \log\frac{n_{j\ldots \ell m}}{n_{j\ldots \ell}}\right) \quad (5.4)$$

with tabulated chi-square values.

But the criteria require ${}_K I_Q$, not ${}_{Q-1}I_Q$, and so an extension of the test is required. Denote by $\phi_{K,Q}$ the likelihood ratio under the null hypothesis, H_K, to that under the new alternate hypothesis, the chain is Q dependent, H_Q, $Q>K$. Then, according to Gates and Tong(1976),

$$\phi_{K,Q} = \phi_{K,K+1}\,\phi_{K+1,K+2}\cdots\phi_{Q-1,Q}$$

and

$${}_K I_Q = -2\log\phi_{K,K+1} - 2\log\phi_{K+1,K+2} \cdots - 2\log\phi_{Q-1,Q} . \quad (5.5)$$

Good (1955) showed ${}_K I_Q$ has a chi-squared distribution with $(S^Q - S^K)(S-1)$ degrees of freedom under H_K.

The likelihood ratio statistic for the criteria, ${}_K I_Q$, is obtained by evaluating (5.5), where each

$$-2\ \log\phi_{K+\nu,K+\nu+1},\quad \nu = 0,1,\ldots,Q-K-1$$

is calculated using (5.4). Akaike's information criterion is given by (5.2) and the Schwartz Bayesian criterion is evaluated using (5.3).

Since both criteria depend on the asymptotic behaviour of the log likelihood ratio they are inherently large sample procedures. In particular, Chin (1977) suggested sample sizes of at least one-thousand are required for stable estimates of the chain order. Chin noted a tendency for the AIC to misrepresent the chain as one of lower than correct order for short samples. Consequently, sample sizes of at least one-thousand days were used for evaluation of both criteria.

Initially it was intended to attempt to determine the sample size required for stable AIC estimates, possibly resolving the discrepancy between Chin's (1977) requirement of one-thousand days and Gates and Tong's (1976) supposedly stable estimates with only sixty days of data. But Katz (1979b) has shown the AIC estimator proposed by Tong (1975) is inconsistent, with a substantial probability of over-estimating the true chain order (0.135 when the true chain order is 1), and a zero probability of under-estimating the chain order. The inconsistency of the AIC, and the results of Katz's simulations to determine the properties of the criteria for finite samples (Katz, 1979a, 1979b), indicate that the AIC may incorrectly select a second order Markov chain as appropriate when the correct order is one. In such a case, attempting to determine the sample size required for

a stable estimate is meaningless. The second or third order selected with all the data may be incorrect, the lower order given by the AIC for less data may simply be the correct choice and not indicative of an unstable estimate at all. The tendency noted by Chin may be a manifestation of the AIC over-estimating the chain order, and not the result of instability because of the reduced sample size. Nevertheless, enough data were available for large samples and so they were used.

The SBC estimator was used to corroborate the chain order selection by the AIC for the cases studied. Katz (1979b) has shown that the SBC estimator is consistent. But Katz (1979a) noted that for a small (0.1) persistence parameter, $p_{11}-p_{01}$, the SBC estimator has a tendency to under-estimate the chain order, even for large sample sizes.

Both the AIC and the SBC, with their respective tendencies for over-estimation and under-estimation of the Markov chain order were applied to large samples to obtain Markov chain order estimates. The AIC and SBC values in Tables 11, 12, and 13, for chains of order zero to four, are smallest for a first order Markov chain. For the cases studied the criteria agreed that a first order Markov chain was appropriate.

5.4 Independence of Daily Amounts

The assumed independence of daily precipitation amounts during the n-day period is crucial to the theoretical development of the models considered here. The assumption enables the modeling of daily precipitation amount in a relatively straightforward manner. The inclusion of a dependence between daily amounts would require a shift in approach to the problem, for example, the precipitation process might be considered a multi-state Markov chain.

The assumption of conditional independence of daily precipitation amounts is questionable. Persistence is common in meteorological variables and the daily amounts cannot be assumed, a priori, to be conditionally independent. Consequently, a check on the dependence between daily precipitation amounts was necessary.

Since the serial correlation between amounts is expected to decrease with increased time between observed amounts, work was concentrated on the dependence of amounts on consecutive wet days, i.e., the amounts observed when two consecutive days were wet. A lack of dependence between amounts on consecutive wet days was considered sufficient to validate the assumption.

Tukey (1977) and Katz (1977b) recommended that a plot of variable pairs be the first step in attempting to detect dependence between random variates. Cleveland et. al. (1975) and Katz (1977b) pointed out that scatter plots for meteorological variables can be uninformative and possibly

misleading. In particular, the large variability of meteorological variables often makes detection of dependence difficult, and second, the often highly skewed nature of the data—a change in density of points along an axis—makes perception of any relationship difficult.

Correlation analysis is not always appropriate. Correlations measure linear relationships, and so the analysis may not detect other forms of dependence. Also, the testing of correlation statistics can be complicated by inappropriate assumptions.

Tukey (1977) recommended processing the scatter plot as a third alternative. Such a procedure, outlined by Katz (1977b) and combining the approaches of Cleveland et. al. (1975) and Tukey (1977), was used here.

The observed pairs of first and second wet day amounts (X_i, Y_i); $i=1,\ldots,N$ are sorted into ascending order of the precipitation amount X_i on the first wet day. The data were then processed in sliding batches of size r, denoted by

$$B_X(i;r)=\{X_i, X_{i+1}, \ldots, X_{i+r-1}\}$$

and

$$B_Y(i;r)=\{Y_i, Y_{i+1}, \ldots, Y_{i+r-1}\}$$

where the B_X and B_Y are batches of abscissa and ordinate values. Denote by T_0 a statistic calculated for each $B_X(i;r)$ that attempts to locate the middle of the batch. T_1 is a similar statistic for the $B_Y(i;r)$. The graphical display consists of a plot of T_1 against T_0. Katz suggested that T_1 be smoothed by a running mean of size 7 before

plotting. Tukey noted that both coordinates should be smoothed and provided an example of possible difficulties when only one coordinate is smoothed (Tukey, 1977, p. 307).

Two statistics, trimmed means and medians were available for T_1. Each provides a location for the middle of the data, yet is somewhat resistant to the effects of outliers, i.e., variate values differing substantially from the middle values. Only a trimmed mean was used for T_0.

The trimmed mean of an ordered sample $Y_1, Y_2, \ldots, Y_N$ is defined (Katz, 1977b)

$$T_M(\alpha_1, \alpha_2) = \frac{(P_1 Y_{[\alpha_1 N+1]} + Y_{[\alpha_1 N+2]} + \ldots + Y_{N-[\alpha_2 N+1]} + P_2 Y_{N-[\alpha_2 N]}}{N(1 - \alpha_1 - \alpha_2)}$$

where $p_i = 1 + [\alpha_i N] - \alpha_i N$, $i = 1, 2$, and the square brackets denote the greatest integer less than or equal to function. The α_1 (α_2) is the proportion of the sample trimmed from the lower (upper) end of the data.

Two smoothers were programmed for use: a moving cosine bell and a running mean of size l, where l was chosen to be an odd integer.

Linear correlation analysis was used to complement the graphical procedure. Although a zero correlation does not always imply independence, Flueck and Mielke (1975) indicated that a zero linear correlation between gamma variates (assuming that daily precipitation amount is distributed as a gamma variate) implies conditional independence. The major difficulty was that the extreme

skewness of the daily precipitation amounts violated the assumption of normality usually required to test the null hypothesis that the correlation coefficient was zero.

Skees and Shenton (1971) discussed the transformation of highly skewed distributions to near normality. After examination of many tranformations they concluded that no one transformation was completely satisfactory. But the tranformations $y=\log x$ and $y=x^{0.1}$ were found to be reasonable, although sometimes overcorrecting for skewness and kurtosis. These transformations were used in the present study.

Correlations between the first and second day precipitation amounts were calculated using the MIDAS (Fox et. al., 1976) statistical package before and after transformation of the data. The MIDAS statistical package provided critical values for the correlation coefficient under the null hypothesis of zero correlation between the first and second day precipitation amount. The critical values were obtained from

$$\rho=t/\sqrt{t^2+(N-2)}, \qquad (5.6)$$

where t is a t-statistic with N-2 degrees of freedom (Haan., 1977).

Despite the claims by Haan (1977) and Fox et. al.(1976) that the test requires variates with normal parent populations, the critical values obtained with (5.6) are applicable for testing the null hypothesis when the correlation is calculated from the untransformed data. Kendall and

Stuart (1967) have shown that (5.6) is applicable as a distribution-free test of the null hypothesis of zero correlation. The accuracy of the test is better for variates that have near normal distributions, but according to Kendall and Stuart the test is adequate for most practical purposes when N is greater than ten.

A relatively horizontal line on the processed scatter plot and a zero correlation coefficient were taken to be indicative of independence between daily precipitation amounts.

Scatter plots of the day two amount versus day one amount for each case are given in Figures 27 to 32. The data in these figures were obtained by application of ABSTR to the development data for each case month. Each pair of consecutive wet day amounts was plotted, with the exception of two pairs for the Beaverlodge-July case. The two pairs were omitted to permit larger axis scales. Logarithmic axes were used to reduce the skewness and kurtosis of the data, so that the plots would be legible. Conclusions based on interpretation of the raw scatter plots are applicable to the logarithmically transformed data.

The processed scatter plots were obtained by applying the trimmed mean, with $\alpha_1=\alpha_2=0.20$, to the untransformed data pairs to obtain T_0 and T_1 for batches of size fifteen. The cosine bell smoother was then applied, with a size l equal to fifteen for all cases, except the Edmonton-January and Medicine Hat-March cases. An l of eleven was used for the

latter cases. The smoothed line was then overlaid on the raw scatter plot.

Generally, the raw scatter plots would support a claim of independence for the transformed data pairs. However, the summer cases do have a number of points in the upper right portion of the figures that suggest a dependence. To make a definite decision on whether or not the first and second wet-day amounts are dependent, on the basis of the raw scatter plots, would be difficult.

The enhanced scatterplot makes a decision easier, but plotting the smoothed line with logarithmic axes has implications for their interpretation. In Figures 27 to 32 the slope of the trend represents the power *a* in the relation

$$\text{Day Two Amount} = b(\text{Day One Amount})^{a}.$$

No trend means the second amount is not functionally dependent on the first. A trend such as that in Figure 31 implies an almost linear dependence on the first wet day amount.

Figures 27 to 32 show that a trend of increasing second day amounts with increasing first day amounts existed in all cases. A rough eyeball estimate of *a* and *b*, for all cases except March at Medicine Hat, showed *a* to be less than 0.2 and *b* to range from 1 to 3. For Medicine Hat-March, the *a* was approximately 0.6 and *b* was approximately 0.3. The figures show that the assumption of the independence of wet day amounts was compromised.

However, there exists the possiblity that this result

may have occurred by chance, for some or all of the cases. This possiblity was examined by using the result given by Flueck and Mielke (1975) and the distribution-free test on the correlation coefficient.

Table 14 contains the calculated correlations, for the original and transformed data, and the critical values for testing the null hypothesis of zero correlation. The five percent significance level was used.

The most striking result was acceptance of the null hypothesis of zero correlation for the untransformed data of the Medicine Hat-March case, yet the null hypothesis was rejected for the transformed data. Figure 31 certainly suggests a linear dependence between the logarithmically transformed data pairs, but at the same time the small intercept explains the lack of linear dependence between the original data. This case illustrates that linear correlation analysis of transformed data is not entirely satisfactory. Even if the null hypothesis of zero correlation is accepted, all that has been supported is a belief of no correlation between the transformed variates. Nothing can be said specifically about the possible conditional dependence between the original variates (Fox et. al., 1976)

The remaining cases were straightforward. The null hypothesis of zero linear correlation was rejected for the Edmonton-January and Medicine Hat-June cases for both the original and transformed data. The null hypothesis was accepted for the Beaverlodge cases and the Edmonton-June

case. The null hypothesis was rejected for the correlation between the tenth root transformed Beaverlodge-July data, but this result was not considered important in light of the Medicine Hat-March results.

In summary, the processed scatter plots indicated that the first and second wet-day amounts were not functionally independent for any of the cases. The correlation analysis showed that the correlation coefficients were statistically different than zero for the Edmonton-January and Medicine Hat-June cases only. Consequently, using the result stated by Flueck and Mielke, the first and second wet-day amounts were independent for the other cases, provided the amounts were distributed as a gamma variate.

5.5 Dependence of X_t's on Y_{t-1}'s

The Katz model assumes that the distribution of daily precipitation amounts, $F_i(x)$ $i=0,1$, is selected according to $Y_{t-1}=i$. A likelihood-ratio test given by Schickedanz and Krause (1970) was used to determine if observations support the use of $F_i(x)$ with different scale parameters, given that the $F_i(x)$ have a common shape parameter. The basics of the test and estimation of the parameters under the two hypotheses was given in Chapter 4. The log likelihood ratio was given by Schickedanz and Krause to be

$$\log\omega = N\left[\log\Gamma(\eta) - \log\Gamma(\eta') - \eta' \log(1/\lambda)\right] + \eta(N_1 \log 1/\lambda_1 + N_2 \log 1/\lambda_2)$$

$$+ (N_1\overline{\log x_1} + N_2\overline{\log x_2})(\eta' - \eta) + N_1\overline{x_1}(\lambda_1 - \lambda) + N_2\overline{x_2}(\lambda_2 - \lambda)$$

The value $-2\log\omega$ was calculated with GAM2 (Wong, 1980) and compared with a tabulated chi-squared variate with one degree of freedom. When the null hypothesis of equal scale parameters was accepted, the estimates were calculated using the pooled data, (4.5), (4.8), and the common shape parameter η. Scale parameters were estimated using (4.5), (4.9), and the common shape parameter η when the null hypothesis was rejected. The results of this test were given in Chapter 4.

5.6 Dependence of X_t's, T_n's and s

A dependence between the number of wet days and total amount of precipitation might be expected, simply because if there are more wet days more precipitation is expected. But whether or not such an expectation is justified is questionable because a large number of wet days, each contributing a small amount of precipitation, may not give as large a precipitation total as one day with a severe storm.

The assumption required by the TW model, that the total amounts, T_n, be independent of the number of wet days, s, in an n-day period was checked by:

1. plotting T_n versus s for the months considered, and
2. obtaining correlation coefficients between T_n and s.

The T_n and s were calculated for each case month, for every year available, including the independent data. Correlation coefficients and scatter plots of T_n versus s were then obtained.

No attempt was made to determine whether or not the individual daily amounts, X_t, were dependent on s.

The null hypothesis of zero correlation between T_n and s was tested using (5.6). The correlations between the total precipitation in the case month and the number of wet days in the case month were significantly (0.01 level) different than zero for all cases. In all cases, but June at Edmonton, the correlations exceeded 0.60. For June at Edmonton the correlation was 0.49. The scatter plots of T_n versus s, which are not included, clearly indicated a dependence between T_n and s. The results show that a larger total precipitation can be expected when there are more wet days in a month, for the cases examined. More importantly, the result clearly indicates a breakdown in one assumption necessary for calculation of the distribution of the total precipitation in n-days by the TW model.

CHAPTER 6.

The Distributions

6.1 General

In this chapter the theoretical distributions calculated for the six cases are examined. The fit of the theoretical distributions were judged both visually and using the Kolmogorov-Smirnov (K-S) test.

The latter was not appropriate for testing the fit of the theoretical to the observed development distributions because the development data were used to estimate parameters for the theoretical curves. Crutcher's (1975) comments on the conservative nature of the test under these circumstances must be kept in mind. To allow use of the K-S test, the asymptotic critical value given by Crutcher (1975) for a normal distribution was used when testing the fit of the distributions for the number of wet days or total precipitation in the n-day period. This approach was justified, assuming that a sample size of 30 or 31 days was sufficiently large for the asymptotic value to be appropriate, because the theoretical distributions for the number of wet days and the total precipitation in the n-day period, calculated using the models, have been shown to be asymptotically normally distributed (Feller, 1956; Katz, 1977c). Katz (1977c) has shown that the distribution calculated for the maximum precipitation in n-days, using the recurrence relation approach, asymptotically approaches the

Type I extreme value or Gumbel distribution. Consequently, Crutcher's asymptotic critical value for the extreme-value distribution was used for testing the fit of the distributions for maximum daily precipitation amount.

Crutcher provided critical values for the K-S test when the location and scale parameters had been estimated for the theoretical distribution. Although those parameters were not directly estimated in this work, sufficient parameters were estimated to completely specify the distribution. Therefore Crutcher's values were appropriate. The K-S test was applicable with the standard critical values when judging the fit of the theoretical distributions to the independent distributions.

The distributions were also examined for the effects of parameter errors and breakdowns in the assumptions. When a parameter estimate or assumption was thought to have affected the calculated distributions, the estimate or assumption was pointed out.

Finally, the distributions were examined in the light of differences between the observed development and independent data samples. But first a brief discussion of case month selection is given.

The AIC and SBC were applied to each month of the year for each of the three sites. Only those months for which a simple Markov chain was appropriate, according to both criteria, were considered for further modeling. The AIC showed a second or higher order Markov chain was appropriate

for a number of months at each station; the SBC showed each month of the year for each station was a simple Markov chain.

The case months for the three stations were selected from those months for which a simple Markov chain was appropriate because of the author's interests. A summer case at each station was desired because the Alberta climate exhibits a summer maximum in monthly precipitation amount. The Beaverlodge-May case was selected because the precipitation process appeared to be quite stationary for that month and location. No specific reason was used to select the January at Edmonton or Medicine Hat during March cases.

6.2 Case I, Beaverlodge-May

Figures 33 and 34 show the theoretical and observed distributions for the number of wet days during the thirty-one day period for the development and independent data sets.

Both the Katz and TW models reproduced the development distribution for the number of wet days in May. However, the calculated distributions overestimated the independent data distribution between 4 and 13 days. Neither of the distributions calculated using the Katz or TW model were significantly different than the observed distributions, at the 0.05 level.

Although the observed development and independent distributions were slightly different, the difference was

not statistically significant, according to the two-sample K-S test at the 0.05 level. The sample sizes of the development and independent data sets used for the K-S tests were 45 and 20 since there were no months of data missing during the 45 years (1914-1958) of the development sample or the 20 years (1959-1978) of independent data.

The figures suggest that the occurrence of precipitation in May at Beaverlodge is adequately modeled by a simple Markov chain. In this case the use of varying transition probabilities resulted in only minor changes to the distribution obtained using constant transition probabilities. This result is not surprising because the daily transition probabilities were found to be stationary for the month.

Despite the good approximation of the precipitation occurrence process by the simple Markov chain, the models did not provide a good representation of the distribution of maximum daily precipitation in the 31 day period. Figure 35 shows that both models overestimated the distribution of the development data for amounts greater than 8mm. The TW model overestimated the observed distribution the worst, by approximately 0.17 near 11mm and by 0.13 near 18mm. The Katz model, with the Das and Mielke parameters, overestimated the distribution by 0.15 and 0.12 respectively near 10mm, and by 0.06 for amounts in excess of 15mm. Each of the calculated distributions underestimated the probability of a daily amount in excess of 15mm by 0.06 to 0.13.

The Katz model with the Das parameters provided the

best fit for amounts up to 7mm, this was because of the good fit of the gamma distribution with Das parameters to the distribution of observed daily amounts for this case. The TW model provided the poorest fit because of the poor fit of the exponential distribution to the observed distribution of daily precipitation amount (Figure 21).

The Katz model, using the Mielke and Das estimates provided distributions that were nearly identical for amounts larger than 15mm. For amounts less than 15mm the Das distribution, as it should, exceeded the Mielke distribution. The different parameter estimates changed the distribution by approximately 0.05 for amounts smaller than 15mm.

For 45 observations the critical value for the K-S test was calculated to be 0.13. Consequently, the TW and Katz-Das distributions were significantly different (0.05 level) than the observed. The null hypothesis that the Katz-Mielke and the development distributions were the same was accepted.

The independent observed and theoretical distributions shown in Figure 36 were not significantly different. The maximum difference of 0.23, near 9mm, between the observed and Katz-Mielke model did not exceed the critical value of 0.29 for 20 observations at the 0.05 significance level.

The theoretical distributions underestimated the probability of a daily precipitation amount in excess of 20mm for the 1959-1978 period; the TW model by up to 0.10.

The most stiking aspect of Figure 36 was the contrast

of the independent observed distribution with the observed distribution shown in Figure 35 for amounts less than 15mm. The independent distribution was 0.10-0.30 higher than the development distribution in the 0-15mm range. However, the maximum difference of 0.33 at 9.9mm was not large enough to reject the null hypothesis that the distributions were the same, by the two-sample K-S test at the 0.05 level. Although the independent sample was less than half the size of the development sample, the variation between the two distributions was an indication that the distribution of maximum daily amount is subject to a large sampling fluctuation.

The distributions for the total amount of precipitation in May are shown in Figures 37 and 38. The TW model provided the best fit to the development distribution in this case. The Katz model with the Das parameters overestimated the observed curve in the 15-55mm range, and although the TW and Katz model with Mielke parameters provided essentially the same distributions for amounts up to 45mm the TW model provided a better fit for amounts greater than 45mm. The distributions calculated using the Katz model with the Das parameters deviated most from the development distribution, by up to 0.10 in the 20-30mm range and near 50mm. This maximum absolute deviation was less than the K-S critical value and so all the theoretical distributions were accepted to be the same as the observed. A 0.05 significance level was used.

The independent distribution in Figure 38 was underestimated by the calculated distributions in the 0-50mm range, and overestimated in the 50-155mm range. The theoretical distributions all underestimated the probability of a monthly precipitation total in excess of 55mm for the 1959-1978 time period.

None of the theoretical distributions were significantly different than the independent distribution for the total precipitation in May at Beaverlodge. The largest K-S statistic for the three curves had a value of 0.21, which was smaller than the critical value of 0.29 at the 0.05 level of significance.

The maximum difference of 0.222, between the development and independent distributions at 24.6mm, was an indication that the distribution of total precipitation was also subject to a large sampling fluctuation. A most important difference between the distributions is the larger precipitation amounts that were observed during the 1959-1978 period. Even if 45 years of data were used to obtain a model that fit the development data very well, the natural variability in the process would not be reflected in the calculated distribution. In this case the probability of a monthly precipitation total in excess of 100mm would be underestimated by five to ten percent. For this case the models provided an adequate representation for the distributions of the number of wet days and the total precipitation amount during the month. The calculated distributions for

maximum daily precipitation fit the two samples poorly, but were between the two observed distributions.

6.3 Case II, Beaverlodge-July

Despite the nonstationarity of the daily transition probabilities that was found by Anderson and Goodman's test for this case, the Katz and TW models resulted in essentially the same distributions. The Katz distribution exceeded the TW distribution by approximately 0.02 near 14mm, not a significant difference. The distributions are shown in Figures 39 and 40.

The theoretical distributions were good approximations to the observed development distribution, particularly over the 8-14 day range. However, in this case the models underestimated the probability of only 4-8 days with precipitation, and underestimated the probability of more than 15 days precipitation. Figure 33 shows a similar feature, although it was less noticeable in the May case.

The models provided a reasonable approximation to the independent data distribution shown in Figure 40. However, the fit was not as good as for the development distribution. The maximum difference of 0.17 between the independent observed and theoretical distributions was not large enough to reject the hypothesis that the distributions were the same.

The maximum difference between the development and independent distributions was 0.16, at 10 days. The

difference was not large enough to reject, by the two-sample K-S test, the null hypothesis that the distributions were from the same parent population.

In this case the simple Markov chain adequately modeled the daily occurrence of precipitation.

The theoretical and observed distributions for the maximum daily precipitation in July during the development period (1914-1958) and independent period (1959-1978) are shown in Figures 41 and 42. None of the calculated distributions fit the observed distributions well. The theoretical distributions did not even give the shape of the observed curves, exhibiting far more curvature than the observed distributions.

The three theoretical distributions were essentially the same for amounts less than 12mm and greater than 48mm. The Katz model distributions, with the Das and Mielke parameter estimates, were the same over the entire range of amounts observed. The underestimation of the probability of amounts in excess of 12mm and 16mm for the development and independent cases was likely the result of the gamma and exponential distributions underestimating the probability of large daily amounts of precipitation.

According to the K-S test, with Crutcher's critical values, the observed development distribution was significantly different than all three theoretical distributions. The K-S statistics were all in excess of the critical value of 0.13, for a 0.05 level of significance.

The observed distribution for the independent sample was similar in shape to the one for the development data. The independent data distribution values were larger than the development values, but the two observed distributions were not different, according to the two-sample K-S test applied with a 0.05 level of significance.

The models did not adequately represent the distribution of the maximum daily amount of precipitation in July at Beaverlodge. All of the models underestimated the probability of a precipitation amount in excess of 20mm by 10% to 15%.

The theoretical and observed distributions for the total amount of precipitation in July at Beaverlodge are shown in Figures 43 and 44. The models provided distributions for the total amount of precipitation that were better approximations to those observed than they did for the distribution of the maximum daily amount of precipitation. But again, the three theoretical curves underestimated the probability of a large precipitation total in the development data. For monthly totals under 80mm the Katz model with the Das parameters provided the best approximation to the development distribution; for totals greater than 80mm, the worst.

The models' underestimation of the observed distribution for amounts less than 30mm for the development data and 50mm for the independent data was a combined effect of two factors. First, the Markov chain underestimated the

probability of less than 4-8 and 4-12 wet days for the development and independent data. Second, the gamma and exponential distributions slightly underestimated the probability of a small amount of daily precipitation. The size of the influences of the two factors was not ascertained, although it may be significant that the independent sample distribution was underestimated more for small amounts than the development distribution when a similar feature was exhibited by the distribution for the number of wet days in the month. The fewer wet days in the independent data resulted in smaller precipitation totals.

The Markov chain's underestimation of the probability of more than 14 wet days in the development sample may be responsible for the underestimation of monthly totals in excess of 80mm, in the same sample. The fact that neither feature was evident in the independent sample is noteworthy. The two previous observations are reasonable because of the correlation between monthly precipitation totals and the number of wet days in the month that was discussed in Chapter 5.

These results suggest that a proper modeling of the number of wet days in the month may be more important than an exact modeling of the daily distribution of precipitation amount. The distribution calculated with the smaller Das parameters was 0.05-0.10 higher than that calculated with the Mielke parameter estimates, for monthly totals of 20-100mm. The increase was similar to that observed in the

May case. Yet the increase in the distribution of the number of wet days in July for less than 12 wet days, from the 1914-1958 period to the 1959-1978 period, may have resulted in an increase of 0.10-0.15 in the distribution for the total amount of precipitation, at amounts ranging from 0-50mm.

The K-S test, using Crutcher's critical values, accepted the hypothesis that the two Katz modeled distributions were the same as the development distribution. The maximum difference of 0.14, between the TW distribution and the observed development distribution at 58mm, was large enough to reject the hypothesis that those distributions were the same.

The null hypothesis that the independent observed and theoretical distributions were the same was not rejected. The K-S statistic had a maximum value of 0.25, the difference between the TW and observed distributions shown in Figure 44. The maximum difference of 0.15 between the observed development and independent samples was small enough that the hypothesis that the two distributions were from the same parent population could not be rejected by the two-sample K-S test.

The models provided distributions that adequately represented the observed distributions for the number of wet days and total amount of precipitation in July at Beaverlodge. The modeled distributions for the maximum daily precipitation in July at Beaverlodge were inadequate.

6.4 Case III, Edmonton-January

Figures 45 and 46 show the distributions for the number of wet days in January at Edmonton for the 1883-1932 development and 1933-1978 independent samples. There were no months of data missing from the samples, so the sample sizes were 50 and 46 respectively.

The Katz and TW models produced identical distributions for the number of wet days in the 31 day period. This result was not surprising because Anderson and Goodman's test showed that the variation in the transition probabilities was not statistically significant, and the same initial probability of a wet day on 31 December was used in both models.

The fit of the calculated distributions to those observed was not good; indeed, the fit to the independent distribution was terrible. The calculated distributions underestimated the number of occurrences of less than 7 wet days in the month and more than 7 wet days in the month's development sample by up to 0.13 and 0.09 respectively. The maximum difference between the calculated and development distributions was just small enough that the hypothesis that the distributions were the same was accepted at the 0.05 level. The critical value used, 0.13, was calculated using Crutcher's (1975) asymptotic values for the normal distribution.

The theoretical curves did not fit the independent data distribution at all well. The models badly overestimated

the distribution for the range of the number of wet days observed. The models underestimated the probability of more than 16 wet days in January by up to 0.18; the maximum difference was near 10 wet days where the theoretical curves exceeded the observed distribution by 0.52. The maximum difference was large enough to reject the null hypothesis that the theoretical and observed distributions were the same.

The distributions from the development and independent data were significantly different. The maximum difference of 0.46 was large enough to reject the null hypothesis that the distributions were from the same parent population, at the 0.05 level using the two-sample K-S test. In this case the difference between the development and independent distributions made it difficult to obtain a calculated distribution with a good fit to both observed distributions.

The surprising aspect of the observed distribution shown in Figure 46 was that the curve suggests that the 1933-1978 period was wetter than the 1883-1932 development period, i.e., the probability of more wet days was higher in the latter period. This result contradicts the earlier result that the probability of a wet day on seven days in January had decreased from the 1883-1937 level to the level observed for the 1938-1978 period. The latter result demonstrates the difficulties that can be encountered when attempting to interpret changes in the probability of a wet day, on 31 consecutive days, in terms of how the entire

period will change, i.e., more or fewer wet days during the period.

Despite the inability of the Markov chain model to adequately represent the number of wet days in the independent sample for January at Edmonton, the models provided a reasonable approximation to the observed distribution for the maximum daily precipitation amount in January during the independent period. The maximum deviation of the theoretical curves from the observed was 0.14, near 10mm. At 10mm the probability given by each of the three distributions was nearly equal, and so none of the three theoretical distributions was significantly different than the observed distribution at the 0.05 level of significance. The distributions are shown in Figure 47.

A possible explanation of the reasonable fit for the maximum daily amount when the fit for the number of wet days was poor is that there were more wet days in the 1933-1978 period, but the amounts on those wet days were smaller than during the 1883-1932 period. This is consistent with the observed trend toward smaller 10-year-mean-daily amounts on a wet day that was found for January at Edmonton.

The TW distribution for the maximum daily amount was essentially the same as the one calculated with the Katz model and the Thom and Greenwood-Durand (TGD) parameter estimates. Use of the Das parameters increased the distribution by up to 0.10 in the 4-6mm range and negligibly downward for amounts greater than 10mm. The Katz model,

with the Das parameters provided the best fit to the distribution of independent data. This model provided the better fit in the 0-10mm range and was marginally poorer for greater than 10mm of precipitation.

Figure 48 shows the distribution of the maximum daily amount in January for the development period. The Katz model with the Das parameters again provided the best fit. This model followed the observed distribution closely up to 7mm and for amounts greater than 17mm. The TW and Katz-TGD models provided the best fit in the 7-11mm range only. Each of the theoretical models underestimated the probability of a maximum daily amount in the 8-16mm range. The maximum deviation of the theoretical curves from the observed was 0.15, near 12mm. This value exceeded Crutcher's critical value of 0.13, for the extreme value distribution, and so the null hypothesis that the theoretical and observed distributions were the same was rejected.

The maximum deviation between the two observed distributions occured near 12mm. The deviation was not large enough to reject the null hypothesis that the distributions were from the same parent population.

Figures 49 and 50 show the distributions for the total amount of precipitation in January. The theoretical distributions calculated using the Katz-TGD and TW models were identical. Use of the Das parameters increased the distribution values above those from the two other models, by up to 0.18.

The Katz-Das distribution gave the best fit to the development distribution for small precipitation totals. But over the entire range of amounts observed the TW and Katz-TGD models provided distributions that fit the best. The maximum deviation of the Katz-Das model from the development distribution was 0.23, so the hypothesis that the Katz-Das model was the same as the observed was rejected at the 0.05 level. The hypothesis that the TW and Katz-TGD modeled distributions were the same as the observed was accepted at the 0.05 significance level. The maximum difference of 0.10 near 5mm did not exceed Crutcher's critical value of 0.13.

The Katz-Das model badly overestimated the independent distribution. The TW and Katz-TGD models provided the best fit to the independent sample distribution, but also overestimated the distribution throughout the range of amounts observed. The null hypothesis that the distributions were the same was rejected when the Katz-Das and independent distributions were compared, but accepted when the TW or Katz-TGD distributions was compared with the distribution from the independent sample.

The hypothesis that there were more wet days with smaller amounts in the 1933-1978 period is not contradicted by the distribution for the total monthly amount of precipitation observed in the independent sample. The larger number of wet days in the independent sample resulted in a downward shift in the distribution for the total monthly

amount and the shift was somewhat compensated for by the trend toward smaller wet day amounts.

Despite the downward shift in the distribution of total precipitation amount, from the development to independent period, the distributions were not significantly different at the 0.05 level, according to the two-sample K-S test.

The Markov chain model did not adequately model the occurrence of precipitation in January at Edmonton, it badly underestimated the variance of the development sample. The sampling fluctuation between the development and independent samples was so large that the modeled distributions were very poor approximations to the independent distribution. However, the Katz model with Das parameters adequately modeled the distribution of maximum daily precipitation in the month. In the troublesome range of amounts from 8-16mm the modeled distribution was between the two observed distributions. The TW and Katz model with the Mielke estimates provided an adequate representation for the distribution of the total amount of precipitation in January at Edmonton.

6.5 Case IV, Edmonton-June

Figures 51 and 52 show the distributions for the number of wet days in June at Edmonton, for the 1883-1932 development sample and the 1933-1978 independent sample.

The TW modeled distribution fits the observed development and independent distributions quite well. The maximum deviation of 0.05 near 7 days in the first instance was not

significant. Neither was the maximum deviation of less than 0.05 near 17 days in the latter case. The two observed distributions were not significantly different.

The Katz model, for the first time, provided a distribution that was appreciably different than that given by the TW model. The Katz model overestimated the observed development and TW distributions by up to 0.22 and 0.15 respectively. According to the K-S test the Katz distribution was significantly different than the distribution for the development data, at the 0.05 level. However, the Katz distribution was not significantly different than the independent data distribution; the maximum deviation between the two was 0.18.

The Katz model overestimated the observed distribution because the Fourier series estimates for the transition probability p_{00} were too large. This problem was noted in Chapter 4. The difference in the distributions can be attributed to the Fourier series estimates for the transition probability because the same initial probability was used for each model, and the models calculated identical distributions when the same transition probabilities were input.

Figure 53 shows the distribution of the maximum daily amount of precipitation in June at Edmonton during the development period. In general, the theoretical distributions were higher than the observed distributions. The Katz model with the Das parameters overestimated the observed

distribution by approximately 0.10 over the middle of the range of amounts observed, and provided the closest fit for amounts greater than 22mm. The TW and Katz model with TGD parameters provided the best fits for amounts less than 18mm.

The maximum deviation of the Katz-Das distribution from the observed development distribution was 0.12—insufficient to reject the null hypothesis that the distributions were the same, at the 0.05 level of significance. Similarly, the Katz-TGD distribution was not significantly different from the observed development distribution, but the TW distribution was found to be significantly different than the observed, at the 0.05 level of significance.

The Katz-Das distribution provided the best fit to the independent sample distribution shown in Figure 54. The three theoretical distributions fit equally well over the 0-25mm range; the Katz-TGD and TW overestimated the observed distribution at larger amounts more than the Katz-Das distribution. None of the theoretical distributions were significantly different than the distribution from the independent sample, at the 0.05 level of significance.

The maximum difference of 0.15 between the two observed distributions, near 8mm, was not large enough to reject the hypothesis that the distributions came from the same parent population, according to the two-sample K-S test applied at the 0.05 level.

The theoretical and development data distributions for

the total amount of precipitation in June are shown in Figure 55. In this case the TW model provided a distribution which was slightly better than that given by the Katz model with the TGD parameters. Neither the TW nor the Katz-TGD distributions were significantly different than the development data distribution. The Katz-Das distribution was significantly different; it badly overestimated the observed distribution for the range of amounts recorded.

An attempt to determine the extent to which the Fourier series estimates for the transition probabilities influenced the models was made. Distributions were calculated using the TGD gamma parameters, and both fixed and varying transition probabilities. Comparison of the distributions indicated that essentially all of the difference between the Katz-TGD distribution and TW distribution could be attributed to the difference in transition probabilities. The difference between the Katz-TGD and Katz-Das distributions resulted from the use of different gamma distribution parameters.

None of the theoretical distributions fit the independent distribution well. But only the TW distribution was found to be significantly (0.05) different than the observed. The critical value was just exceeded near 42mm. The Katz-Das distribution was best for amounts up to 45mm; the observed distribution then followed the TW distribution for amounts greater than 90mm.

The maximum difference between the two observed

distributions was 0.16, at 42mm. This difference was not large enough to reject the null hypothesis that the distributions were from the same parent population, according to the two-sample K-S test at the 0.05 significance level.

Although the TW model provided a reasonable approximation to both the development and independent distributions for the number of wet days in June, the distributions calculated for the maximum daily amount in June were not adequate because of the consistent underestimation of the observed distributions for amounts greater than 20mm. Although the TW model provided a good approximation to the development distribution for total precipitation, the model providing the overall best fitting distributions was the Katz-Mielke. This model was chosen because it seemed to provide the distribution giving the best fit to both the development and independent distributions.

6.6 Case V, Medicine Hat-March

Figures 57 and 58 show the distributions for the number of wet days in March at Medicine Hat for the 1884-1933 development and 1934-1978 independent samples. Two months of data, 1886 and 1887, were missing from the development sample so the sample size was 48. The independent sample, with 45 months of data, was complete.

The Katz and TW distributions were similar in this case; the Katz was 0.05 higher near 5 wet days. The difference was attributed to the fact that the Fourier

series estimates for the transition probabilities differed from the mean transition probabilities for March. In the Medicine Hat cases the mean probability of a wet day for the case month was used for the initial probability, rather than the probability of a wet day on the day previous to the month.

There was little basis on which to choose which model provided the better fit in the development case. Both models underestimated the probability of less than 0-4 wet days and overestimated the probability of less than 6-11 wet days. A similar feature appeared in cases I, II, and III; it suggests the Markov chain has underestimated the variance of the precipitation occurrence process.

The maximum deviation of the TW distribution from the observed was 0.12, less than the 0.15 difference between the observed and Katz distributions at 8 days. Using Crutcher's critical values, only the Katz distribution was significantly different than the observed, at the 0.05 level of significance.

The theoretical distributions were poor approximations to the distribution of the number of wet days in March obtained from the independent sample. Both models overestimated the distribution over the range in the number of wet days observed, i.e., they underestimated the probability of more than N wet days, for N in the range 0-14. The theoretical distributions deviated from the observed by more than 0.3 at 6 wet days; the null hypothesis that the

distributions were the same was rejected for both models.

The hypothesis that the two observed samples were from the same parent population was also rejected. The maximum difference of 0.32 was sufficiently large to reject the hypothesis at the 0.05 level of significance by the two-sample K-S test.

The models did not provide satisfactory distributions for the maximum daily precipitation in March at Medicine Hat. In the development case, shown in Figure 59, the Katz-Das distribution fit reasonably well for low and high precipitation amounts. However, it underestimated the probability of daily amounts of 3-15mm. The maximum deviation of 0.14 was just large enough to reject the null hypothesis that the distributions were the same.

The Katz-TGD and TW distributions differed by only a small amount. The difference of up to 0.15 between these models and the Katz-Das model resulted from the use of different gamma distribution parameters.

The TW and Katz-TGD distributions underestimated the probability of maximum daily amounts less than 6mm, and underestimated the probability of daily amounts in excess of 7mm. The maximum deviation of these theoretical curves from the observed (0.10) allowed acceptance of the null hypothesis that the theoretical and observed distributions were the same.

The fit of the theoretical distributions to the independent distribution, shown in Figure 60, was worse than

that for the development sample. The Katz-Das model overestimated the observed distribution for the entire range of amounts observed. The maximum deviation of 0.21,near 10mm, was significant (0.05 level). The Katz-TGD and TW distributions fit reasonably well up to 6mm, but underestimated the probability of maximum daily amounts in excess of 7-32mm by up to 0.20. The maximum difference was just small enough to allow acceptance of the hypothesis that the independent observed and the Katz-TGD or TW distribution, or both, were the same. The K-S test was applied with a 0.05 significance level.

The two observed distributions had the same shape in this case, but the development distribution was approximately 0.10 higher for the range of amounts observed. The null hypothesis that the distributions were from the same parent population was accepted at the 0.05 level, by the two-sample K-S test.

The observed distributions for the total precipitation in March are shown in Figures 61 and 62, for the development and independent samples. The TW and Katz-TGD models again provided the best approximations to the observed development distribution, but none of the theoretical distributions fit the independent distribution properly.

The TW and Katz-TGD models underestimated the development distribution for small amounts (less than 15mm) and overestimated the distribution for the larger amounts observed. This feature suggests that those models did not

account for the entire variability that was observed in the total amount of precipitation in March. The maximum deviation of the TW modeled distribution from the development distribution was 0.14, near 8mm. The maximum difference between the Katz-TGD distribution and the observed was also at 8mm, but was only 0.10. Crutcher's critical value at the 0.05 level was 0.13, and so the null hypothesis that the development and theoretical distributions were the same was rejected for the TW modeled distribution, but accepted for the Katz-TGD.

The small difference (0.05) between the TW and Katz-TGD distribution was attributed to different transition probabilities and different gamma distribution parameters. The large amount (up to 0.23) by which the Katz-Das distribution deviated from the Katz-TGD distribution provided an example of how the distribution can fluctuate in response to changes in the gamma distribution parameters.

All three theoretical distributions badly overestimated the distribution from the independent observed data. Each of the three were significantly different than the observed distribution, at the 0.05 level.

The maximum difference of 0.27 near 15mm, between the development and independent sample distributions, was not large enough to reject, by the two-sample K-S test, the null hypothesis that the distributions were from the same parent population. But the large sampling fluctuation indicated that efforts to obtain models which fit the observed

distributions of samples of size 50 to within 10% may not be worthwhile. The two-sample K-S test indicated that observed distributions fluctuating from the first by up to 0.27 would be accepted as being from the same parent population. Can one expect to model the ensemble any better than this when the sample size is limited?

6.7 Case VI, Medicine Hat-June

Figures 63 and 64 show the distributions for the number of wet days in June at Medicine Hat for the 1884-1933 development and 1934-1978 independent samples. No data were missing during these time periods, so the sample sizes were 50 and 45 respectively.

The models did not provide distributions that fit the observed distributions well. Katz's recurrence relation approximated the development distribution quite well, up to 6 days. For more than 6 wet days neither model provided a distribution that was a particularly good approximation. The Katz model overestimated the observed distribution by up to 0.15 at 11 wet days. The TW model provided a distribution that underestimated the observed up to 8 wet days, then overestimated the observed for more than 8 wet days. In terms of absolute deviation from the observed, the TW model provided the best fit to the development distribution with a maximum difference of 0.11 between the two.

According to the K-S test, using Crutcher's critical value of 0.13, the Katz distribution was significantly

different than the observed development distribution; the TW distribution was not.

The difference between the Katz and TW distributions was attributed to the differing transition probability estimates. The Katz distribution was nearly 0.10 higher than the TW distribution because the Fourier series gave p_{00}s that were larger than the monthly mean value for that transition probability.

The relationship of the TW distribution to the observed suggests that the model underestimated the variability in the number of wet days observed during June at Medicine Hat.

The models provided distributions with the same shape as the distribution from the independent sample, but the TW and Katz models overestimated the observed distribution by up to 0.17 and 0.25 near 10 days. The TW distribution was not significantly different (0.05 level) than the observed, in contrast to the Katz distribution.

The maximum difference of 0.21 between the two observed distributions at 6 days was not large enough to reject the null hypothesis that the distributions came from the same parent population.

The distributions of the maximum daily precipitation in June at Medicine Hat are shown in Figures 65 and 66. The three distributions fit the development distribution equally poorly, and fit the independent distribution, from 0-28mm, equally well.

The model distributions fit the observed development

distribution well for small and large daily amounts, but underestimated the probability of a daily amount in excess of 15-35mm.

The maximum deviation of the three theoretical distributions from the development distribution was near 18mm, and exceeded Crutcher's critical value in each case. Therefore, the theoretical distributions were considered to be significantly different than the observed.

The TW and Katz-TGD models fit the independent distribution very well up to 20mm, and underestimated the probability of daily totals in excess 20mm. The TW model underestimated the probability of a maximum daily amount in excess of 40mm by 0.15, but the difference was not significant according to the K-S test. The Katz-TGD and Katz-Das distributions exceeded the observed by 0.14 and 0.11 near 40mm.

The two observed distributions were not significantly different, according to the two-sample K-S test with a 0.05 level of significance.

The Katz-TGD model provided the best-fitting representation of the development distribution for the total amount of precipitation in June. The distributions are shown in Figure 67. The maximum deviation of the Katz-TGD distribution from the observed was 0.10, insufficient to reject by the K-S test the null hypothesis that the distributions were the same. The TW distribution also fit well, but did not follow the observed as closely as the Katz-TGD

for amounts up to 70mm. The TW distribution was also accepted to be the same as the observed, by the K-S test.

The Katz-Das model badly overestimated the observed distribution. This distribution was significantly different than the observed; the difference between the Katz-Das and the Katz-TGD can be attributed to the different gamma distribution parameters.

The Katz-TGD distribution also fit the independent distribution the best. However, in this instance the model underestimated the variability of the total monthly precipitation amounts that were observed. The model underestimated the probability of smaller amounts and amounts in excess of 70mm. The TW distribution fit poorly, but was not significantly different than the observed. The Katz-Das model badly overestimated the observed distribution, and was significantly different than the observed distribution. The distributions are shown in Figure 68.

The maximum difference between the two observed distributions was 0.11, insufficient to reject the hypothesis that the distributions came from the same parent population.

The TW model provided the best approximation to the number of wet days in June at Medicine Hat, but underestimated the variability of the process. None of the models approximated the distribution for the maximum daily amount of precipitation in June very well, and underestimated the probability of large daily amounts. The Katz model with the TGD parameters provided a good approximation

for the distribution of the total precipitation in June at Medicine Hat, but underestimated the variability of the total precipitation amounts in the independent sample.

CHAPTER 7.

Summary and Suggestions

7.1 The Preliminaries

In this study two stochastic models were used to calculate probability distributions of monthly precipitation characteristics for two months at each of three locations in Alberta. The abilities of the Katz and Todorovic-Woolhiser models to reproduce the time series distributions of the precipitation characteristics were fair, but a number of problems were identified. These problems will be returned to later. The steps leading to application of the models will be summarized first.

The Katz and Todorovic-Woolhiser models used different theoretical approaches to develop equations for the generation of distributions of the number of wet days, the maximum daily precipitation, and the total amount of precipitation during a month. Two computer programs based on the two approaches were written to calculate the distributions for each of the six cases.

The computer routine for the TW model was much faster than the one for the more general Katz model. For example, to calculate the 30 day distribution for the Medicine Hat (June) case the TW routine required less than 1s of computer time while the Katz routine required 14s. The Katz model required 52s of time to evaluate the distribution when the interval size (used in Simpson's rule to evaluate the

convolutions) was halved from 0.5mm to 0.25mm; the required time varied with the square of the number of intervals used.

For the same parameters the computer routines calculated distributions that were identical. The theoretical distributions differed because of parameter differences and not because of differences in approach.

In general, the use of the Fourier series estimates for the transition and initial probabilities did not improve the performance of the Katz model. This result was consistent with the hypothesis that the transition probabilities were stationary for all of the cases considered except July at Beaverlodge, according to Anderson and Goodman's test. Despite Anderson and Goodman's test showing that the daily transition probabilities were nonstationary for July at Beaverlodge, the distributions of the number of wet days during the month calculated by each of the models were nearly the same, and representative of the observed distribution.

For June at Edmonton the overestimation of p_{00} by the Fourier series caused the Katz model to significantly overestimate the observed development distribution. In this case the use of Fourier series estimates for the transition probabilities were not satisfactory. The equal weight given to each raw estimate when calculating the series coefficients was responsible. Perhaps a method in which the raw estimates for the month of interest were more heavily weighted would be useful.

The cumulative periodogram method used for the selection of the number of Fourier series harmonics to include in the series for the initial and transition probabilities was satisfactory.

In retrospect, more importance could have been placed on the selection of the initial probability that was used in the application of the models. The author arbitrarily selected an initial probability for the cases. Although the maximum difference between p and p′ given in Table 1 for the different cases is only 0.063, for June at Edmonton, the differences possibly resulted in significant upward or downward shifts in the calculated distributions. Possibly the stationary Markov chain probabilities π_0 and π_1 (defined in Appendix A) should have been used for the initial probability of the Markov chain, as was done by Katz (1977c) in his earlier work. This would provide the long term probability of an initial wet day to the models. That probability should be representative of the ensemble of time series for the occurrence of precipitation.

Mielke's iterative procedure for estimation of the shape and scale parameters for the gamma distribution worked well for the cases considered. The iterative procedure produced shape and scale parameters that were essentially the same as those calculated using the Thom or Greenwood and Durand procedures. Despite a possible bias in the Mielke estimates, which can be corrected (Haan, 1977, p104), the application of GAM2 is a good method of obtaining shape and

scale parameters for the gamma distribution.

The fit of the gamma and exponential distributions to the observed distributions of daily amount was not good. Generally the observed distributions were underestimated for smaller amounts and overestimated for larger amounts. The exception was for May at Beaverlodge. In that case the gamma distributions with the Das estimates did not underestimate the observed distributions for smaller amounts, but still overestimated the observed distribution at larger amounts. The better approximation at smaller amounts resulted from the inclusion of the number of traces in Das' parameter estimation procedure.

The gamma distribution was not able to assume the shape of the observed distributions of daily precipitation amount. The gamma distribution's overestimation of the observed distribution for large precipitation amounts is believed to be the major reason for the inability of the Katz and TW models to approximate the distributions of maximum daily precipitation amount.

Many assumptions about the precipitation process were made in this study. Some of the assumptions were examined, and the salient points are summarized here. Little evidence can be offered to support the use of Fourier series estimates of the transition probabilities for time periods of one month or less. According to Anderson and Goodman's test the transition probabilities could generally be considered stationary for the cases examined, and even for July at

Beaverlodge for which the transition probabilities were apparently nonstationary, the two Markov chain models produced nearly the same distributions for the number of wet days in the period. This means, not that the TW distributions were unbiased, but that the use of Fourier series did not produce an appreciable improvement in model results when time periods of one month were considered. The use of Fourier series may make the Katz model perform better than the TW model for time periods longer than one month.

The simple two-sample t-test used to detect nonstationarity in the precipitation occurrence process was not entirely satisfactory. The test was incapable of showing whether or not the large number of significant and negative t-statistics was indicative of a continuous downward trend in the probability of precipitation. Interpretation of test results for a series of days, in terms of long term nonstationarity of the dependent Y_t process, was difficult and the author is uncertain of their implications. It is questionable whether or not the applications of the two-sample t-test for consecutive days were independent, and so the supposedly significant number of rejections may not be indicative of nonstationarity in the precipitation occurrence process.

A number of methods were used to examine the stationarity of the mean of the wet-day precipitation amounts. Generally the tests indicated that the X_t process was stationary in the mean, with the exception of a significant

downward trend in the ten-year mean wet-day amounts for January at Edmonton. The influence of this trend could not be identified conclusively in the observed distributions, but possibly explained the reasonable fit of the theoretical distributions of maximum daily and total monthly precipitation to the independently-observed distributions while the Markov chain badly underestimated the number of wet days in the independent sample. This part of the study is left with some reservations. Undoubtedly entire studies have been concerned only with the stationarity of precipitation time series (Potter, 1976)

The models assumed that the Y_t process was a simple Markov chain. This assumption was turned into a selection criterion for the cases studied. The SBC and AIC were both applied to each month of the year for the sites chosen. The cases were selected from those months for which both the SBC and AIC agreed that a first-order Markov chain was appropriate. This method of case selection ensured that a simple Markov chain was appropriate for modeling the precipitation occurrence process.

The models also required the assumption that wet-day amounts were conditionally independent. This assumption was generally found to be somewhat compromised. The graphical display showed what seemed to be a functional dependence between consecutive wet-day amounts. Correlation analysis showed that the consecutive wet-day amounts were conditionally dependent for the January at Edmonton and June at

Medicine Hat cases only. The assumption was not a good one, but it was not strongly violated in all cases.

The final assumption checked was that the total monthly amount of precipitation and the number of wet days in the month were conditionally independent. This assumption was found to be poor for each case considered in this study, and is a significant shortcoming in the theoretical development of the Todorovic and Woolhiser model.

7.2 Application of the Models

Both the TW and Katz models adequately represented the precipitation occurrence process with a first-order Markov chain. For June at Edmonton and Medicine Hat the Todorovic-Woolhiser model performed better than the Katz model, but this was because of an overestimation of p_{00} by the Fourier series and is not an indication of a major flaw in the Katz model. There was little basis on which to choose which model better represented the occurrence of precipitation.

The first-order Markov chain models seemed to underestimate the variability in the precipitation occurrence process for a number of cases. This appeared in the calculated distributions by a relative underestimation of the probability of both a large and small number of wet days with respect to the observed distributions. This problem was most pronounced when the theoretical distributions for January at Edmonton were compared with the distribution from the development sample for that case.

The case studies showed that the accurate modeling of both the development and independent samples that was hoped for would not be achieved. Undoubtedly the authors initial expectations were too high, but he must question the utility of modeling distributions from samples that have large sampling fluctuations. The sampling fluctuation was such that the maximum absolute difference between the observed distributions exceeded 0.15 in four of the six cases examined, and exceeded 0.3 in two of the six. The two larger deviations, for the January at Edmonton and March at Medicine Hat cases, were large enough that the two sample K-S test rejected (at the 0.05 level) the hypothesis that the two samples were from the same parent population. This suggested that attempts to model distributions to within 0.05 (say) would require much larger samples than the approximately 50 years that was used in this study.

The two models seemed least able to cope with the distribution of maximum daily precipitation during the month. Generally the models underestimated the probability of large maximum daily precipitation totals for the case months considered. This problem is believed to be because of the underestimation of the probability of large daily precipitation amounts by the gamma and exponential distributions.

The Katz model, using the more general gamma distribution, provided better approximations to the distributions of maximum daily precipitation amount. However, neither the Mielke, TGD, or Das parameters could be considered to

provide consistently better distributions.

The models' abilities were mediocre at best. A distribution fitting the daily precipitation amounts better than either the exponential or gamma distributions, is required, particularly for large daily amounts.

In a few cases, particularly Beaverlodge (July), the theoretical distributions underestimated the probability of both large and small maximum daily precipitation amounts for the month. Although this feature was not as common for the maximum daily precipitation as it was for the number of wet days in the month, it suggested that the models underestimated the variability of the maximum daily precipitation during a month.

For a few cases the maximum deviation between the observed development and independent distributions exceeded 0.15. It is apparent that larger samples are required to develop distributions that model the ensemble of monthly maximum daily precipitation amount.

The Katz model with the TGD parameter estimates for the gamma distribution generally provided the best approximations to the distributions of total monthly precipitation for the cases considered. The Katz model with Das parameters overestimated the observed distributions for both the Edmonton and Medicine Hat cases, but did provide a good approximation for the Beaverlodge July case. The TW model frequently calculated distributions that were nearly the same as those provided by the Katz model with the TGD

parameters.

With the exception of March at Medicine Hat the TW and Katz-TGD distributions adequately approximated the observed distributions, but did not approximate the distributions as closely as had been hoped for. In a number of instances (January-Edmonton, June-Medicine Hat) the TW and Katz-TGD distributions were between the observed development and independent distributions. The large fluctuation in the March samples for Medicine Hat was responsible for the inadequacy of the models in that case. Although the TW and Katz-TGD adequately fit the development distribution, a downward shift in the distribution for the independent period (maximum deviation 0.27) caused the theoretical distributions to badly overestimate the observed distribution.

In a few cases the theoretical models appeared to underestimate the variability that was observed in the total monthly precipitation. The notable instances were for the independent distributions of May at Beaverlodge, June at Edmonton, and June at Medicine Hat, and the development distributions for July at Beaverlodge and March at Medicine Hat.

7.3 Suggestions for Further Work

During the latter stages of this work it was apparent that a few statistics summarizing the distributions would have been useful. In particular, it would be useful to append a subroutine that calculated the mean, mode and

variance of the distribution to the computer models. This would enable a quantitative comparison of the theoretical and observed distributions, in addition to the subjective comparison that was done in this study.

Overall, the Katz model is probably the better model, not necessarily because of its performance in this study, but because of its potential. Todorovic and Woolhiser (1975) suggested that a model using a distribution more general than the exponential would be worthwhile and the Katz model is such a model. However, in further work it would be necessary to attempt to obtain distributions that model the daily amount of precipitation better than the gamma distribution. Such distributions might be the Pareto or Bessel distributions. Time constraints made pursuit of this objective impossible for this work.

The sampling fluctuation of the precipitation characteristics examined in this study was large enough to be of concern. Although in many instances the fluctuation was not so large that the distributions had to be considered to be from different parent populations, the fluctuation was large enough with samples of approximately 50 in size that the distributions of the development and independent samples appeared markedly different. Further work to determine whether or not these large fluctuations are stochastic in nature or if they are the result of inhomogeneity or nonstationarity of the time series is necessary. Such work might determine the sample size required to achieve specified

confidence limits for the occurrence, the maximum daily, and the total amount of precipitation during an n-day period.

Tables

TABLE 1. Initial and Transition Probabilities for the Todorovic and Woolhiser Model

Station	Beaverlodge		Edmonton		Medicine Hat	
Case	May	July	January	June	March	June
P_{01}	0.196	0.270	0.186	0.346	0.162	0.241
P_{11}	0.470	0.523	0.426	0.541	0.309	0.442
p	0.269	0.361	0.245	0.428	0.190	0.301
p'	0.216	0.390	0.219	0.365	0.195	0.282
d	0.274	0.253	0.240	0.196	0.147	0.201

TABLE 2. Fourier Series Coefficients for the Initial and Transition Probabilities

Station	Beaverlodge					
Harmonic	1-P		P_{10}		P_{00}	
	A	B	A	B	A	B
0	0.699	–	0.508	–	0.780	–
1	0.025	0.023	–	–	0.027	0.017
2	-0.033	-0.017	–	–	-0.030	-0.012
3	0.022	-0.010	–	–	0.019	-0.009
4	-0.008	0.020	–	–	-0.012	0.015

Station	Edmonton					
Harmonic	1-P		P_{10}		P_{00}	
	A	B	A	B	A	B
0	0.735	–	0.560	–	0.795	–
1	0.090	0.014	0.046	0.003	0.078	0.014
2	-0.066	-0.029	-0.046	-0.008	-0.054	-0.028
3	0.014	-0.004	–	–	–	–
4	0.002	0.020	–	–	–	–
5	0.005	-0.021	–	–	–	–

Station	Medicine Hat					
Harmonic	1-P		P_{10}		P_{00}	
	A	B	A	B	A	B
0	0.798	–	0.625	–	0.842	–
1	0.031	-0.020	0.027	0.007	0.023	-0.019
2	-0.035	0.001	-0.025	0.030	-0.028	-0.007
3	0.006	-0.026	0.007	-0.044	–	–
4	–	–	0.025	0.019	–	–

TABLE 3. Mielke Parameters for the Gamma Distribution

Station / Case	Beaverlodge May	Beaverlodge July	Edmonton January	Edmonton June	Medicine Hat March	Medicine Hat June
η	0.825	0.803	1.014	0.819	0.985	0.877
λ_0	0.216[1]	0.169[3]	0.361	0.128	0.391	0.127
λ_1	0.156[2]	0.127[4]	0.361	0.128	0.391	0.127
α	0.005	0.004	0.110	0.239	0.718	0.591
Cases	375	503	380	642	282	452

[1] K-S statistic 0.097, cases 201
[2] K-S statistic 0.123, cases 174
[3] K-S statistic 0.141, cases 241
[4] K-S statistic 0.087, cases 262

TABLE 4. Das Parameters for the Gamma Distribution at Beaverlodge

	May	July
η_0	0.655[1]	0.592
λ_0	0.192	0.143
K-S	0.037	0.096
Cases	201	241
η_1	0.634[1]	0.721
λ_1	0.134	0.119
K-S	0.068	0.068
Cases	174	262

[1] Observed and theoretical distributions not significantly different at 0.05 level.

TABLE 5. Gamma Distribution Parameters

Station	Edmonton		Medicine Hat	
	January	June	March	June
Thom				
η before				
bias corrected	1.015	0.819	0.986	0.877
η	1.006	0.815	0.976	0.871
λ	0.358	0.127	0.387	0.126
Greenwood and Durand				
η before				
bias corrected	1.014	0.819	0.985	0.877
η	1.006	0.815	0.975	0.871
λ	0.358	0.127	0.387	0.126
K-S	0.111	0.071	0.131	0.080
Das				
η	0.598	0.531	0.535	0.485
λ	0.269	0.0981	0.288	0.0878
K-S	0.115	0.097	0.139	0.158
Cases	380	642	282	452

TABLE 6. Exponential Distribution Parameters

Station	Beaverlodge		Edmonton		Medicine Hat	
Case	May	July	January	June	March	June
λ	0.223	0.179	0.356	0.156	0.397	0.145
K-S	0.148	0.148	0.110	0.113	0.136	0.107
Cases	375	503	380	642	282	452

TABLE 7. Number of Occurrences of Amounts that are Multiples of One-Tenth of an Inch of Precipitation

Station	Beaverlodge			Edmonton		Medicine Hat	
Case	March	May	July	January	June	March	June
2.3mm	8	9	11	3	17	3	7
2.5mm	60	14	17	30	24	32	29
2.8mm	5	9	6	5	16	4	6
4.6mm	–	–	–	–	–	3	–
4.8mm	1	4	12	1	6	–	6
5.1mm	31	8	9	22	10	15	12
5.3mm	2	3	6	4	5	1	4

TABLE 8. Beaverlodge Normals

Total Precipitation (mm) and Number of Wet Days 1941-1970 (Canadian Normals, 1973)

J	F	M	A	M	J	J	A	S	O	N	D
32.0	29.2	23.1	22.1	41.1	61.7	64.3	57.4	38.9	26.4	30.7	27.7
12	11	11	8	10	11	12	11	12	9	11	11

Total Precipitation (mm) 1931-1960, and Number of Wet Days 1941-1960 (Climatic Normals, 1968)

J	F	M	A	M	J	J	A	S	O	N	D
32.0	29.5	25.7	21.1	40.6	56.4	64.0	51.8	40.1	31.8	32.8	29.2
11	12	11	8	9	12	12	11	11	9	10	11

Total Precipitation (mm), 31 years, and Number of Wet Days, 10 years (Climatic Summaries, 1947)

J	F	M	A	M	J	J	A	S	O	N	D
32.3	22.9	29.7	19.8	41.7	53.6	56.1	46.0	43.4	28.2	32.5	30.5
9	12	11	8	10	12	12	12	11	9	11	10

TABLE 9. Edmonton Normals

Total Precipitation (mm) and Number of Wet Days 1941-1970 (Canadian Normals, 1973)

J	F	M	A	M	J	J	A	S	O	N	D
25.1	20.1	16.8	23.4	37.3	74.7	83.8	71.6	35.8	18.5	18.5	21.3
12	10	10	8	9	12	13	12	9	6	9	11

Total Precipitation (mm) 1931-1960, and Number of Wet Days 1941-1960 (Climatic Normals, 1968)

J	F	M	A	M	J	J	A	S	O	N	D
24.1	19.6	21.1	27.9	46.5	80.0	84.8	64.8	34.3	22.9	22.4	25.1
12	10	10	7	9	13	13	12	9	7	8	11

Total Precipitation (mm), 55 years, and Number of Wet Days, 8 years (Climatic Summaries, 1947)

J	F	M	A	M	J	J	A	S	O	N	D
22.4	16.3	19.3	22.4	47.0	77.7	84.3	59.7	33.8	19.1	19.1	20.6
12	9	10	8	12	15	14	12	9	9	11	12

TABLE 10. Medicine Hat Normals

Total Precipitation (mm) and Number of Wet Days 1941-1970 (Canadian Normals, 1973)

J	F	M	A	M	J	J	A	S	O	N	D
22.6	18.3	19.3	25.1	38.1	63.5	38.6	39.4	33.0	17.0	16.3	16.5
9	8	7	6	8	10	8	7	7	5	6	8

Total Precipitation (mm) 1931-1960, and Number of Wet Days, 1941-1960 (Climatic Normals, 1968)

J	F	M	A	M	J	J	A	S	O	N	D
21.6	20.3	24.9	24.9	41.7	58.9	34.5	39.1	37.8	20.6	19.6	19.1
9	9	8	6	8	11	8	8	7	5	7	7

Total Precipitation (mm), 56 years, and Number of Wet Days, 10 years (Climatic Summaries, 1947)

J	F	M	A	M	J	J	A	S	O	N	D
16.0	14.5	16.0	19.6	40.9	61.5	42.7	34.5	28.7	15.7	17.5	17.8
10	10	9	8	10	11	9	7	7	5	7	7

TABLE 11. Beaverlodge Markov Chain Order

	May		July	
K	AIC	SBC	AIC	SBC
0	78.98	0.372	73.93	-4.68
1	-16.76	-90.13	-12.17	-85.54
2	-14.30	-77.19	-10.05	-72.94
3	-8.79	-50.71	-6.98	-48.91
4	0.0	0.0	0.0	0.0

TABLE 12. Edmonton Markov Chain Order

	January		June	
K	AIC	SBC	AIC	SBC
0	77.46	-2.73	36.39	-43.31
1	-3.26	-78.10	-18.80	-93.18
2	-2.25	-66.40	-15.35	-79.11
3	-1.31	-44.08	-8.55	-51.05
4	0.0	0.0	0.0	0.0

TABLE 13. Medicine Hat Markov Chain Order

K	March AIC	March SBC	June AIC	June SBC
0	3.18	-76.40	52.24	-27.46
1	-23.95	-98.22	-4.34	-78.72
2	-21.13	-84.79	-1.26	-65.01
3	-14.28	-56.22	3.54	-38.97
4	0.0	0.0	0.0	0.0

TABLE 14. Correlation between Day One and Day Two Amounts

Case	Transformation x	Transformation log x	Transformation $x^{0.1}$	critical value[1]	pairs
Beaverlodge					
May	0.1213	0.1032	0.1095	0.1488	174
July	0.0875	0.1208	0.1300	0.1212	262
Edmonton					
January	0.1597	0.2805	0.2683	0.1543	162
June	0.0765	0.0920	0.098	0.1061	342
Medicine Hat					
March	0.1115	0.3655	0.3530	0.2108	87
June	0.1808	0.1769	0.1848	0.1388	200

[1] Critical value at 0.05 level of significance.

Figure 1. Location of Alberta stations used in this study.

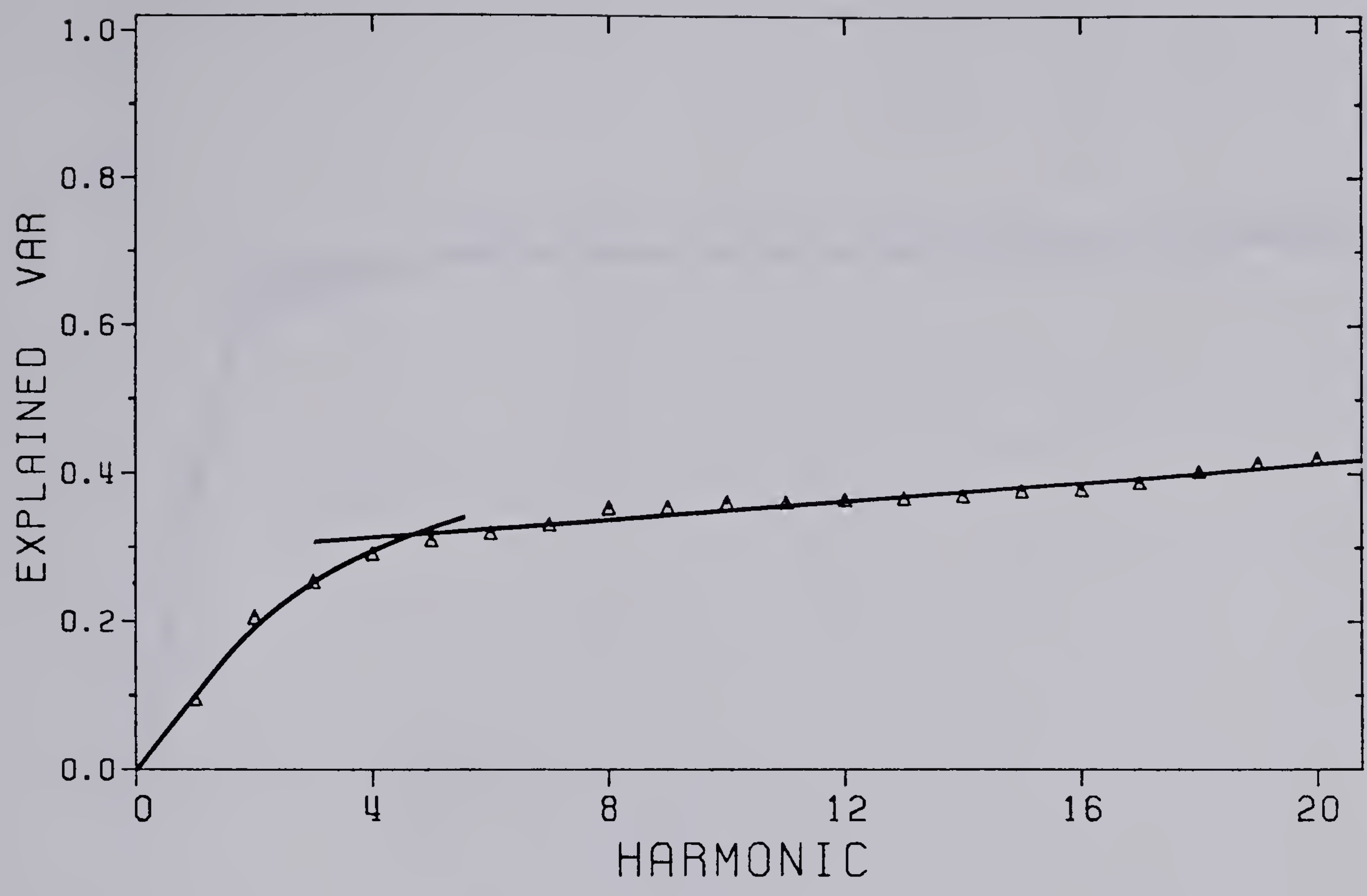


Figure 2. Cumulative periodogram for the probability of a dry day at Beaverlodge.

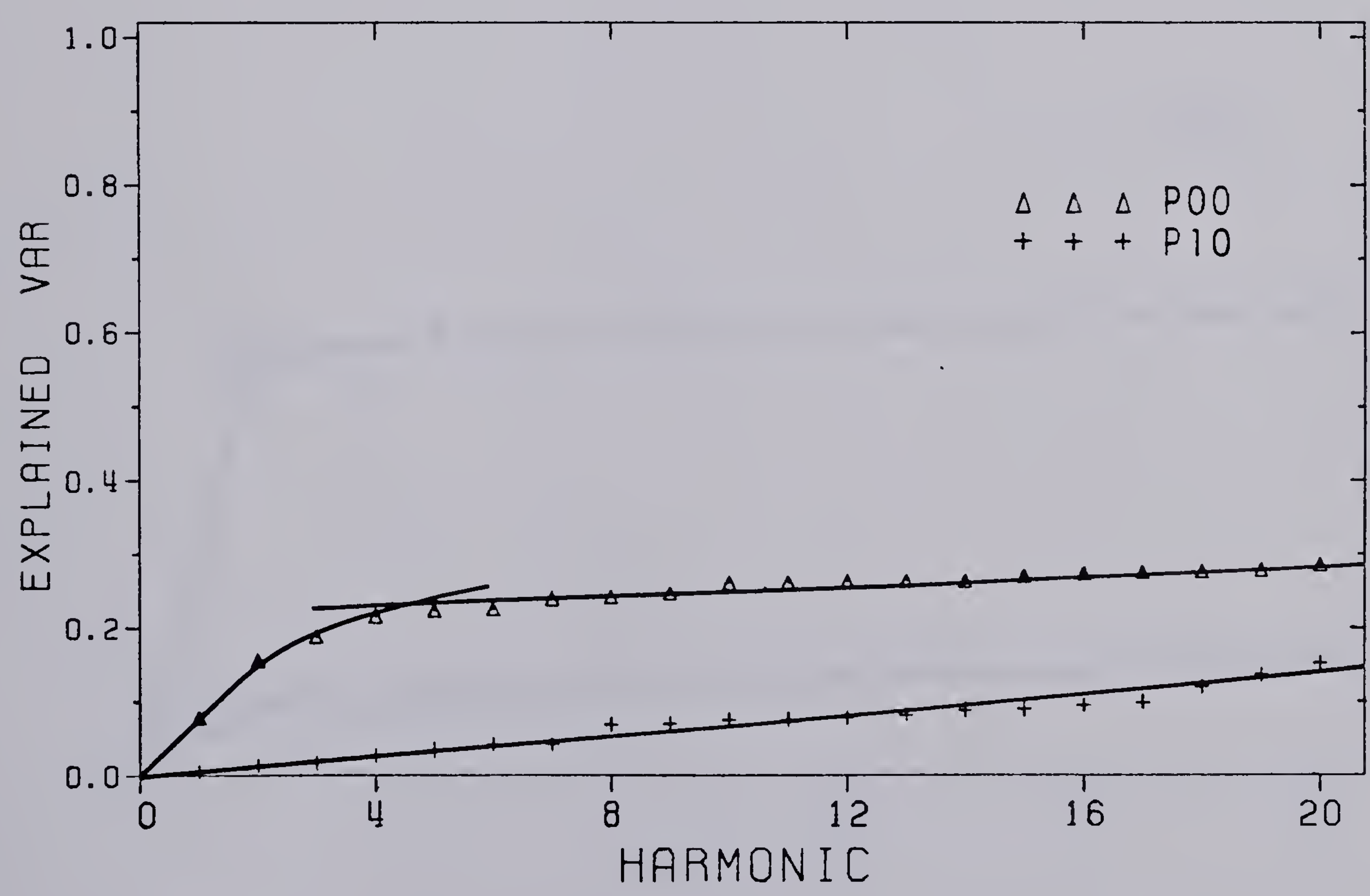


Figure 3. Cumulative periodogram for the transition probabilities at Beaverlodge.

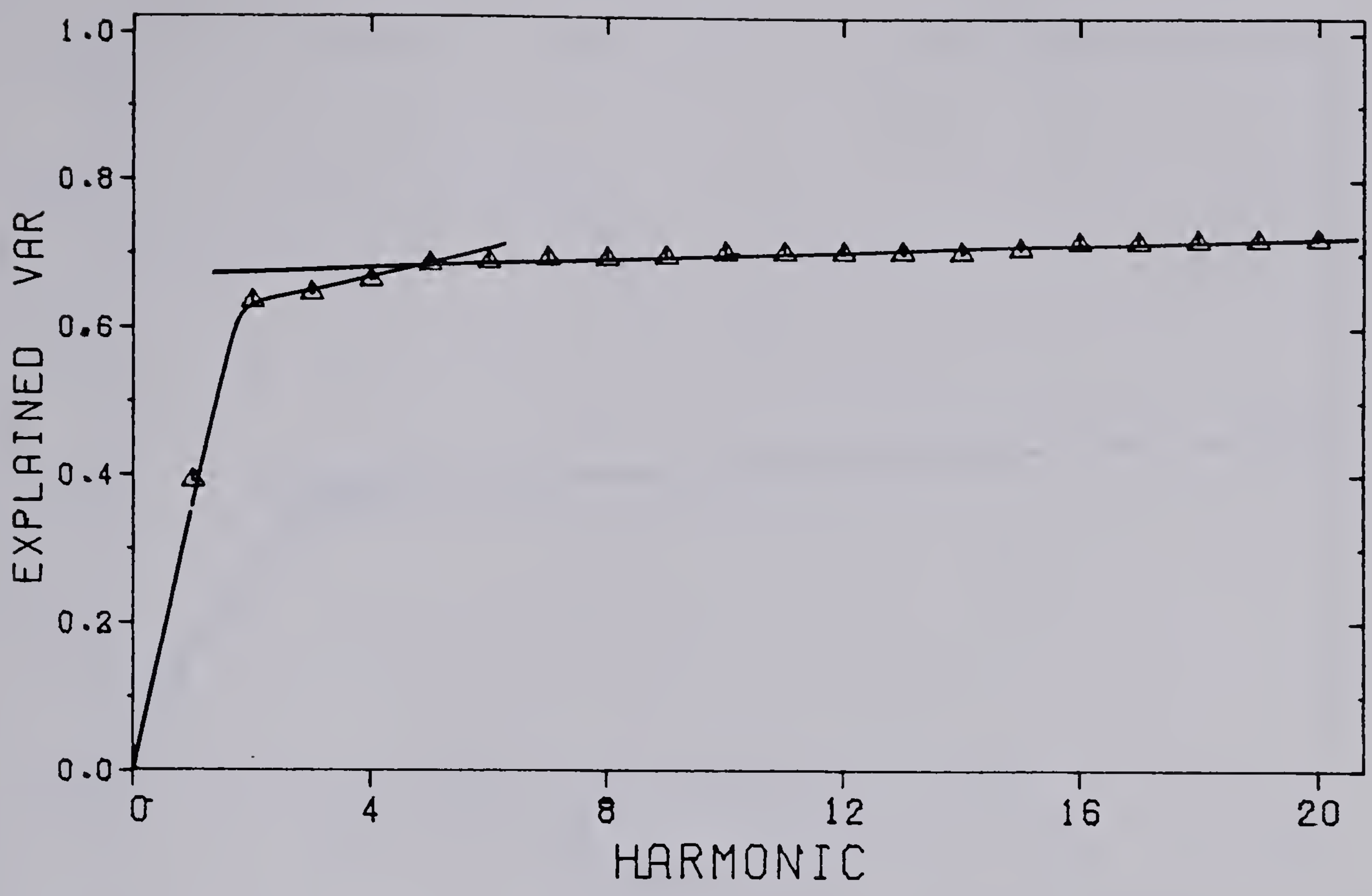


Figure 4. Cumulative periodogram for the probability of a dry day at Edmonton.

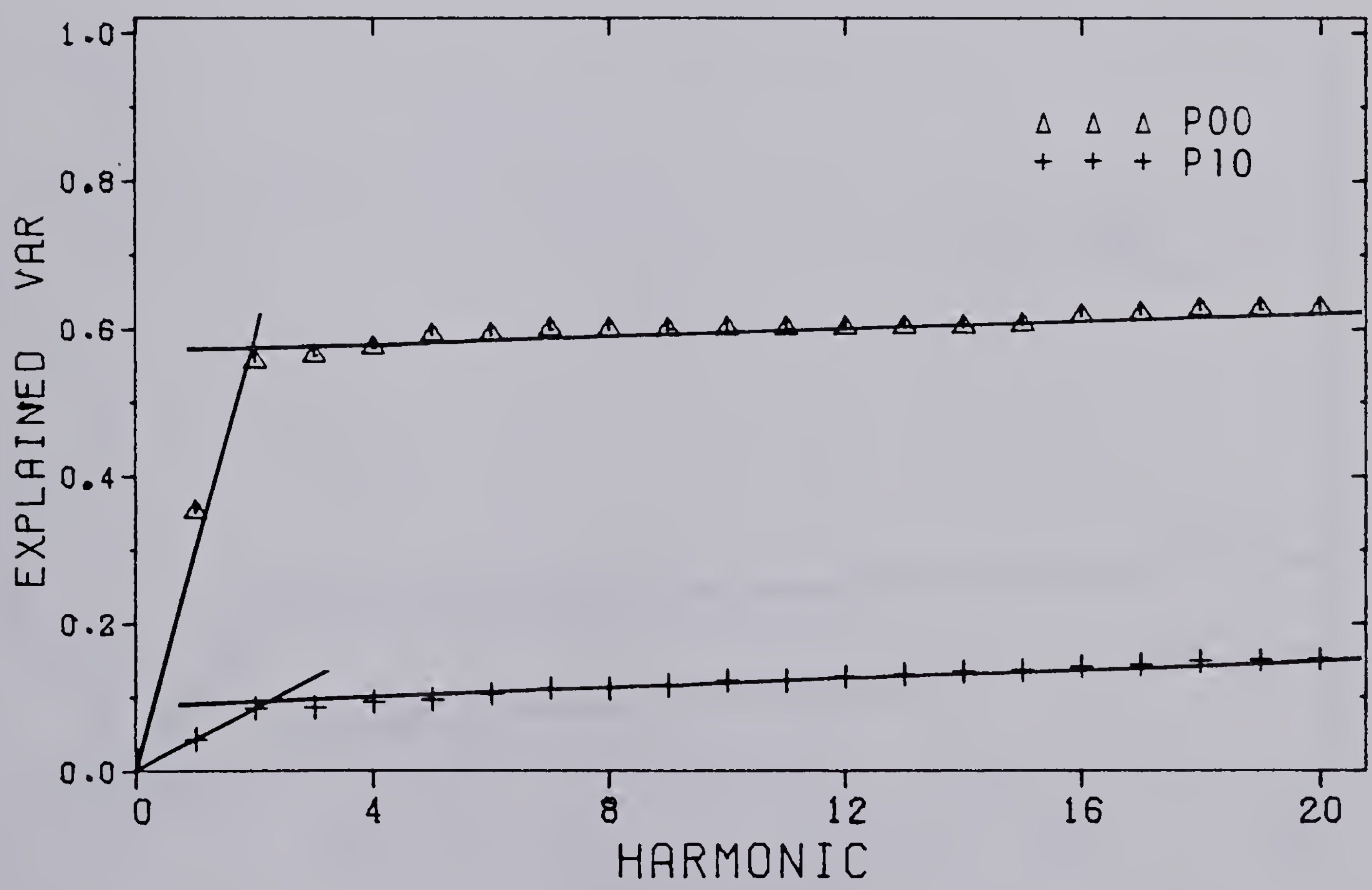


Figure 5. Cumulative periodogram for the transition probabilities at Edmonton.

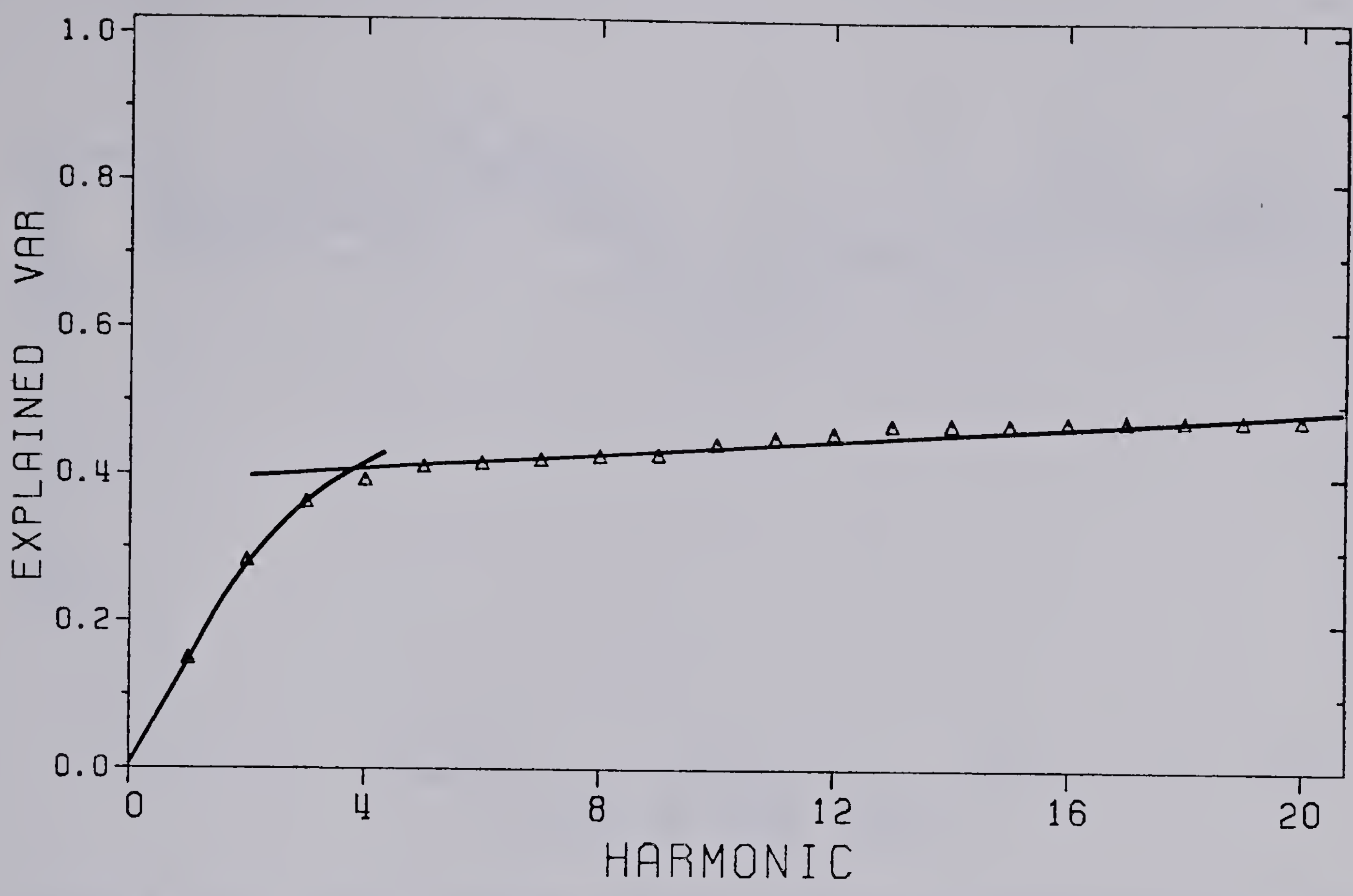


Figure 6. Cumulative periodogram for the probability of a dry day at Medicine Hat.

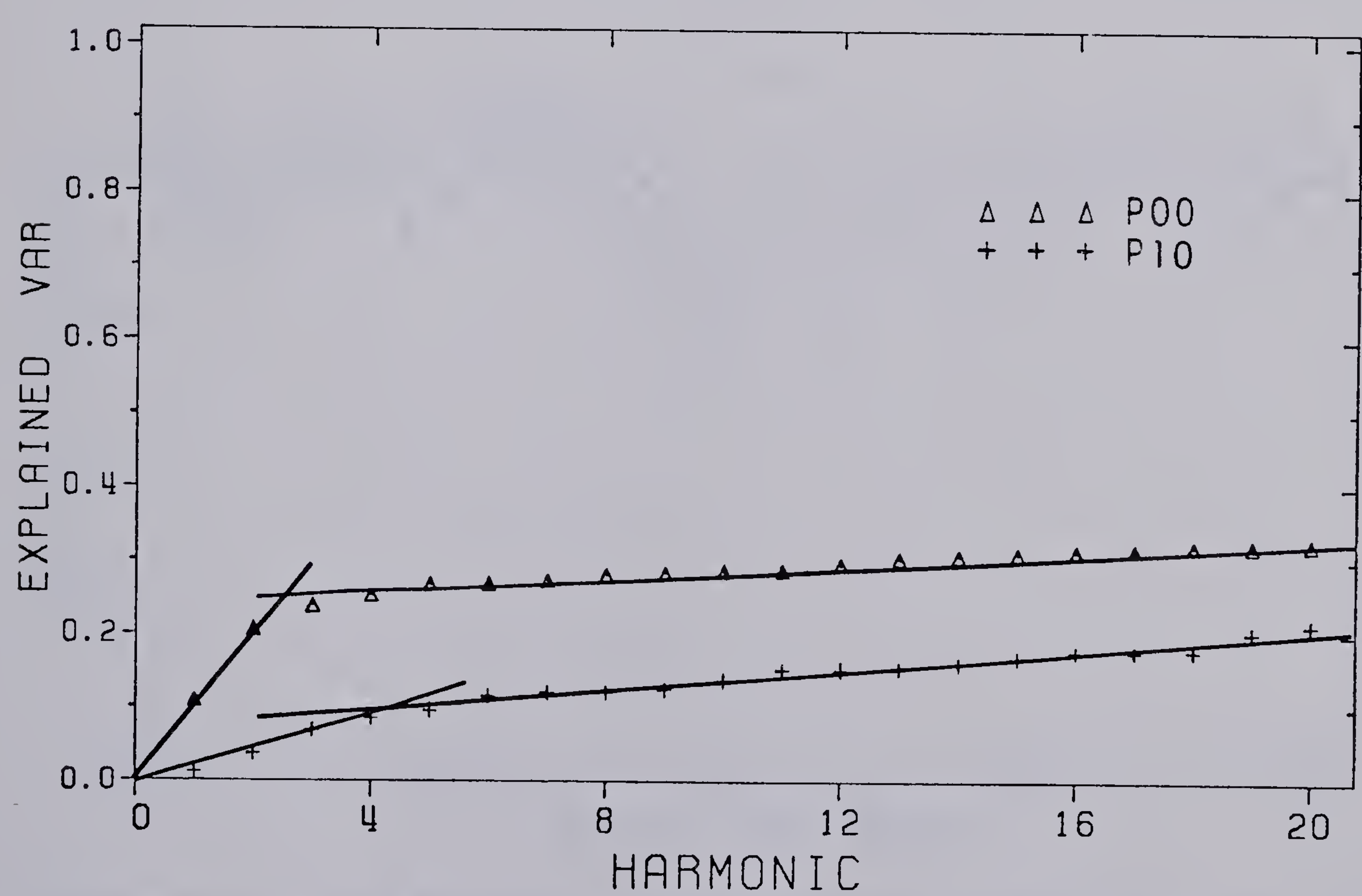


Figure 7. Cumulative periodogram for the transition probabilities at Medicine Hat.

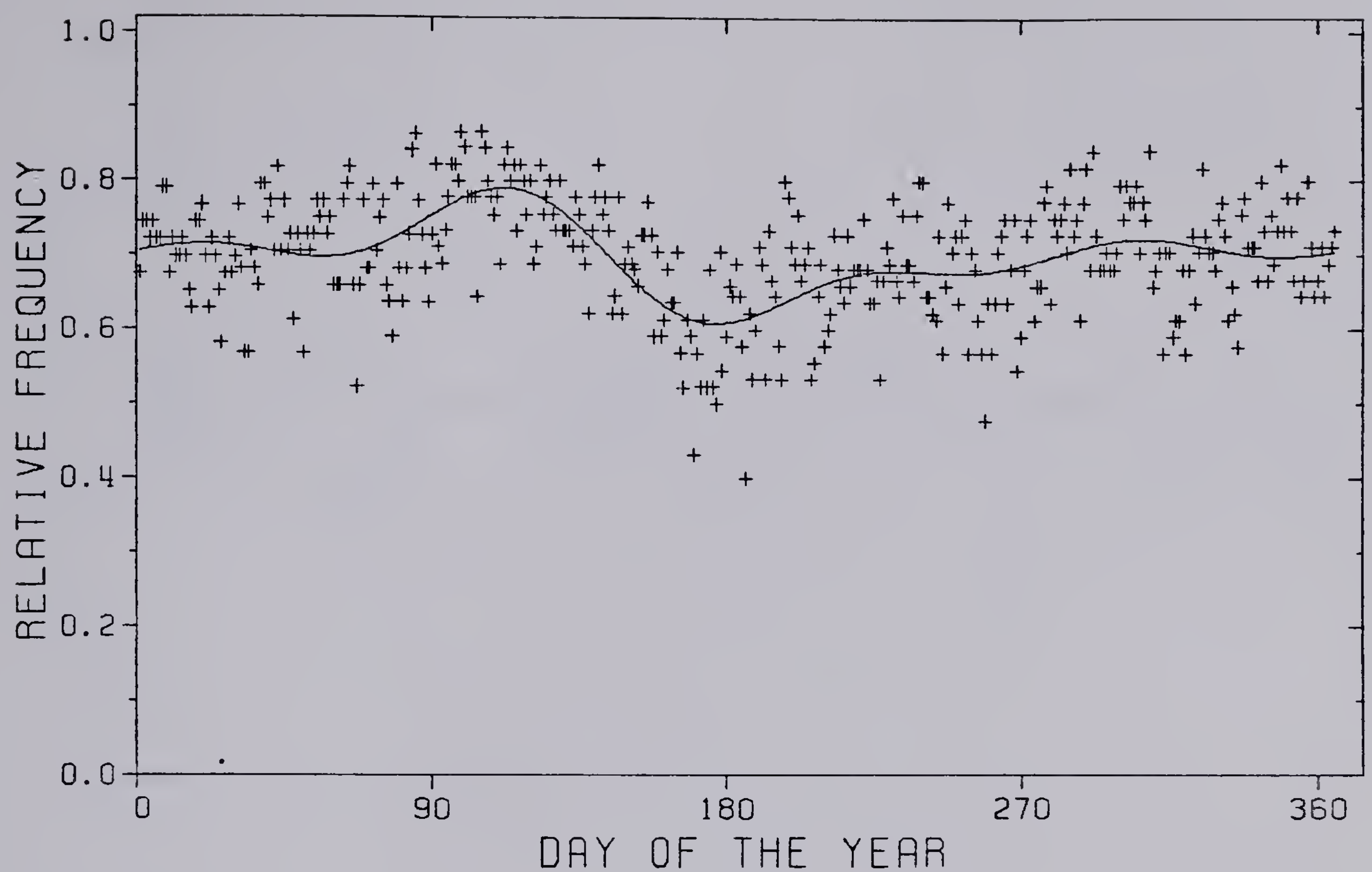


Figure 8. The Fourier series and raw estimates for the probability of a dry day at Beaverlodge throughout the year.

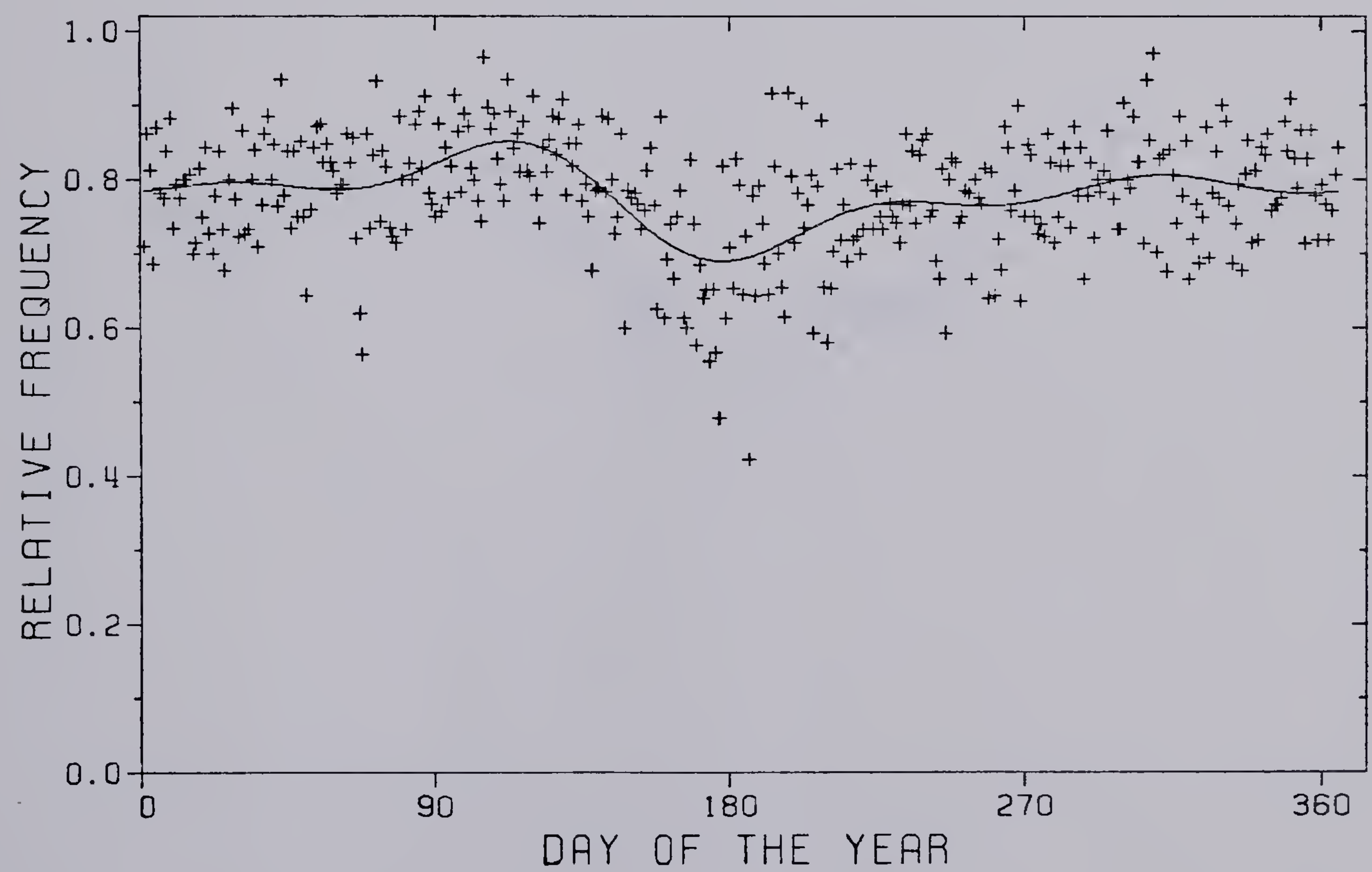


Figure 9. The Fourier series and raw estimates for P00 at Beaverlodge throughout the year.

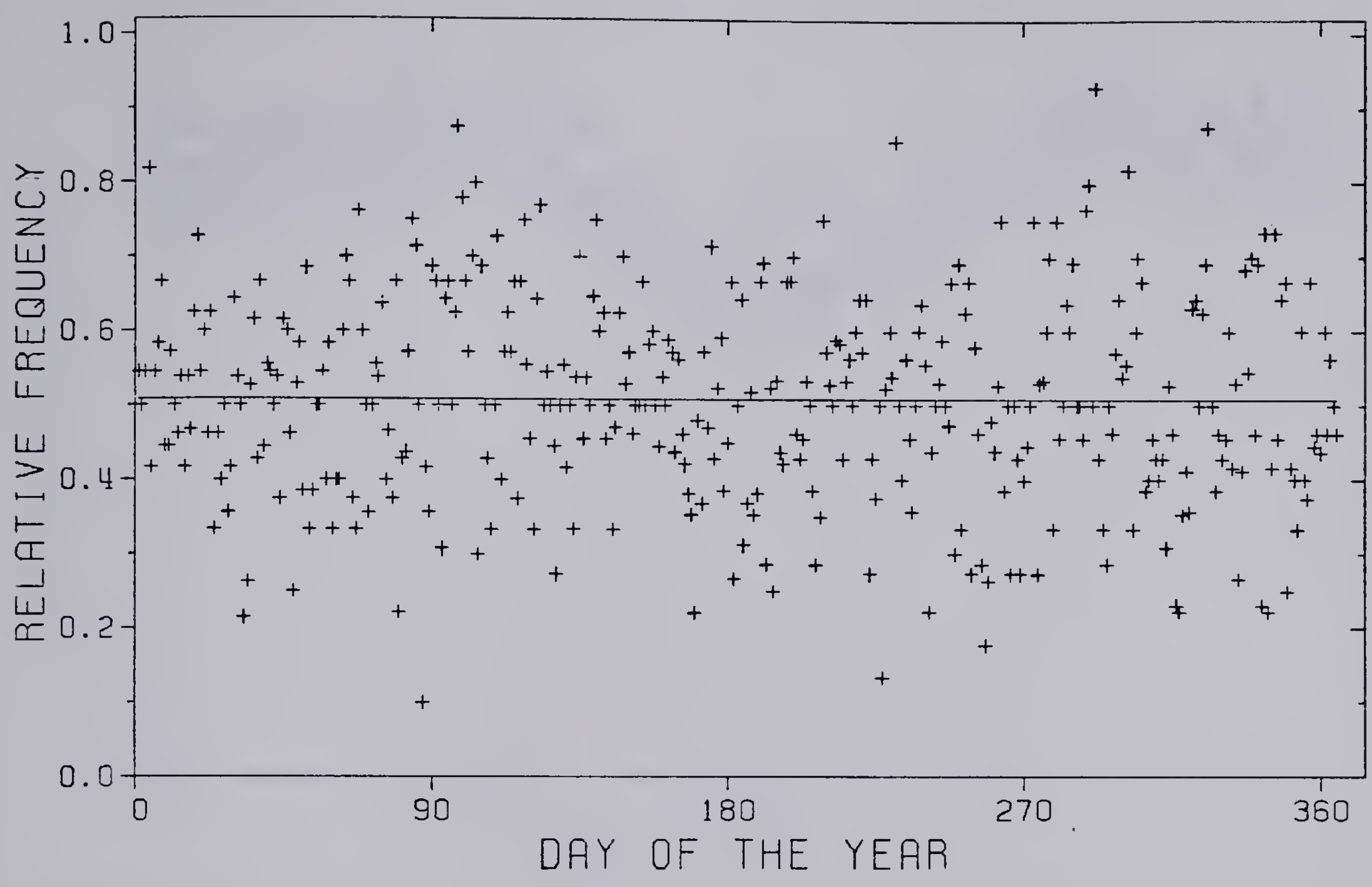


Figure 10. The Fourier series and raw estimates for P10 at Beaverlodge throughout the year.

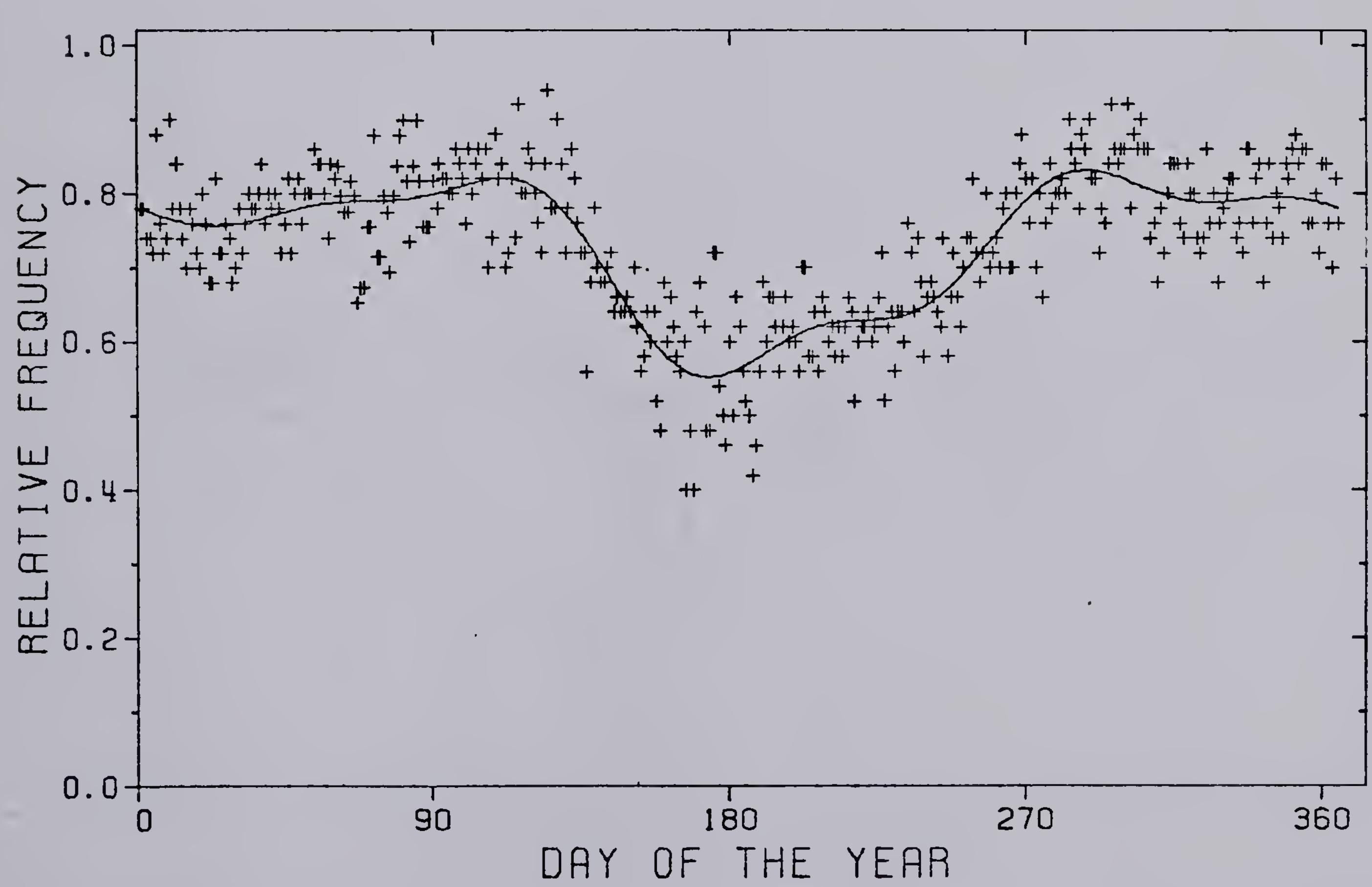


Figure 11. The Fourier series and raw estimates for the probability of a dry day at Edmonton throughout the year.

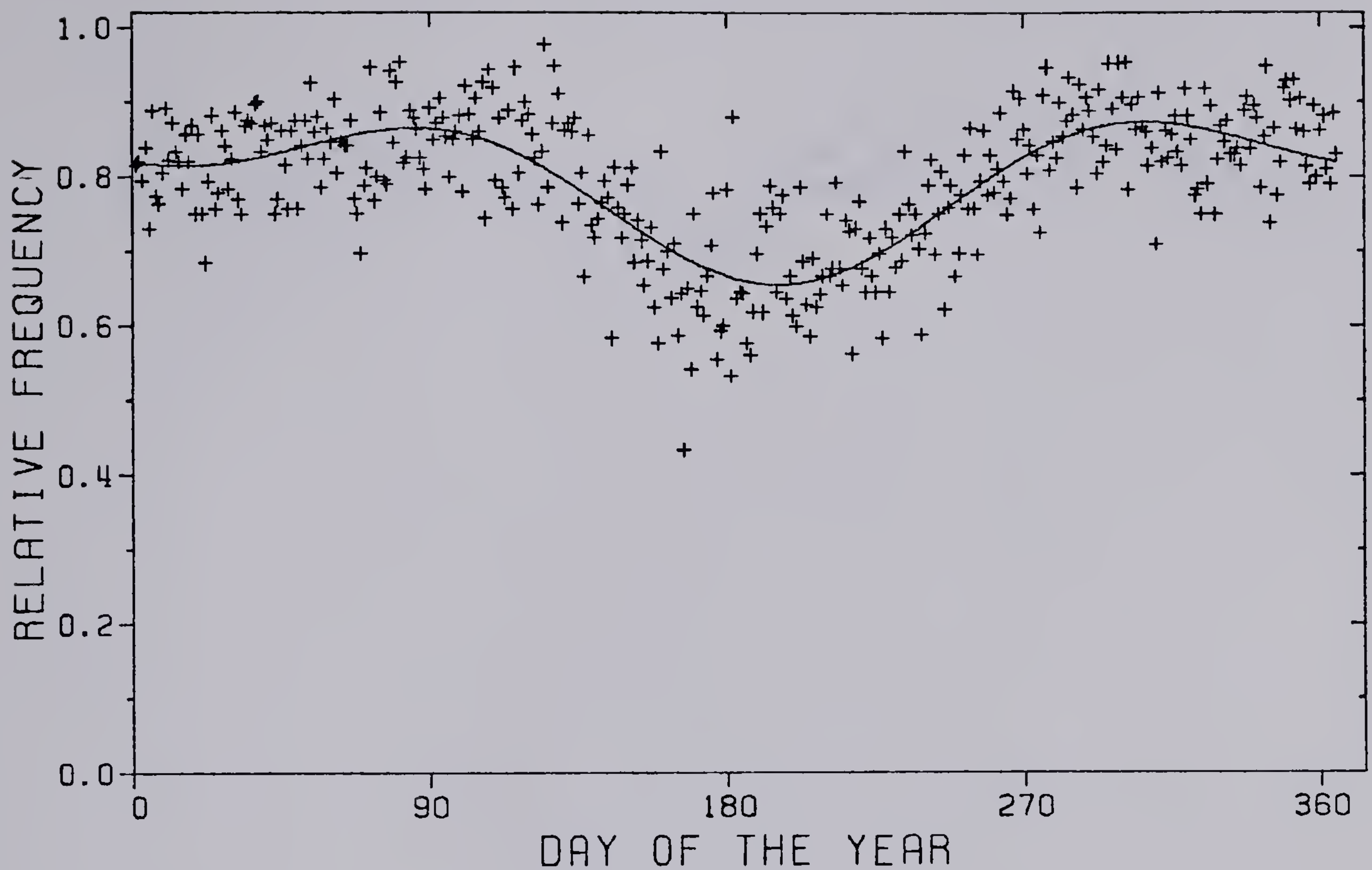


Figure 12. The Fourier series and raw estimates for P00 at Edmonton throughout the year.

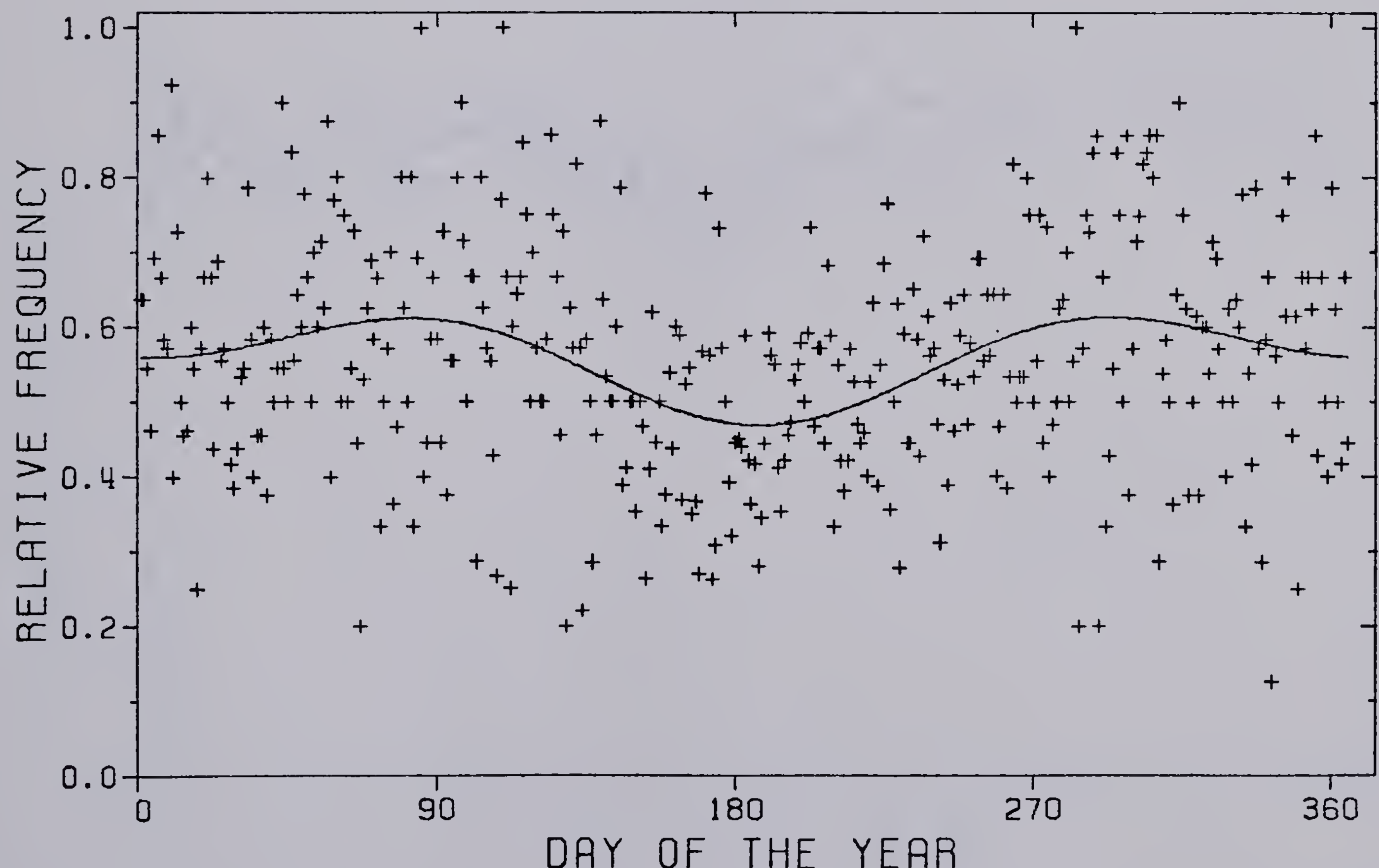


Figure 13. The Fourier series and raw estimates for P10 at Edmonton throughout the year.

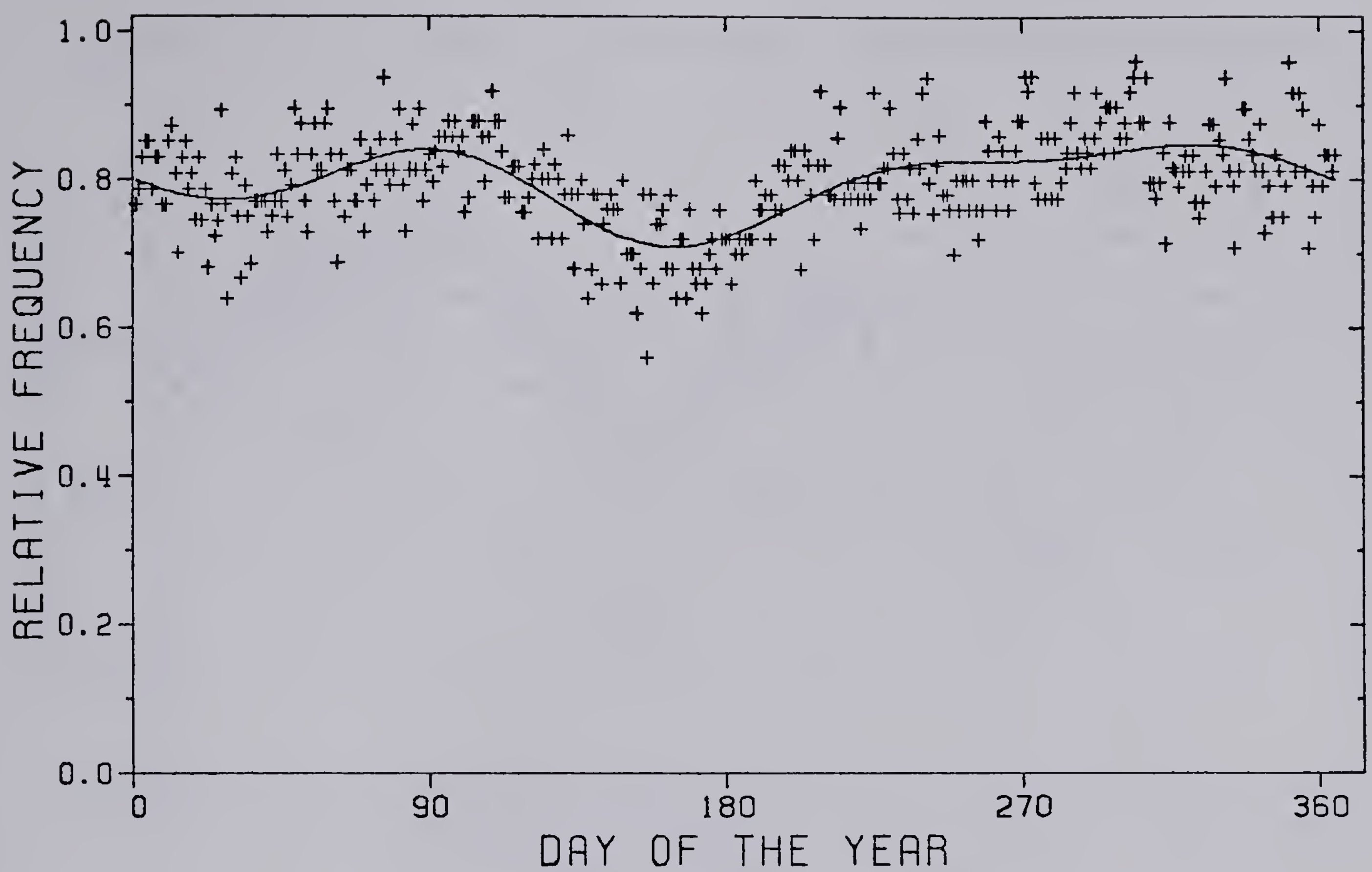


Figure 14. The Fourier series and raw estimates for the probability of a dry day at Medicine Hat throughout the year.

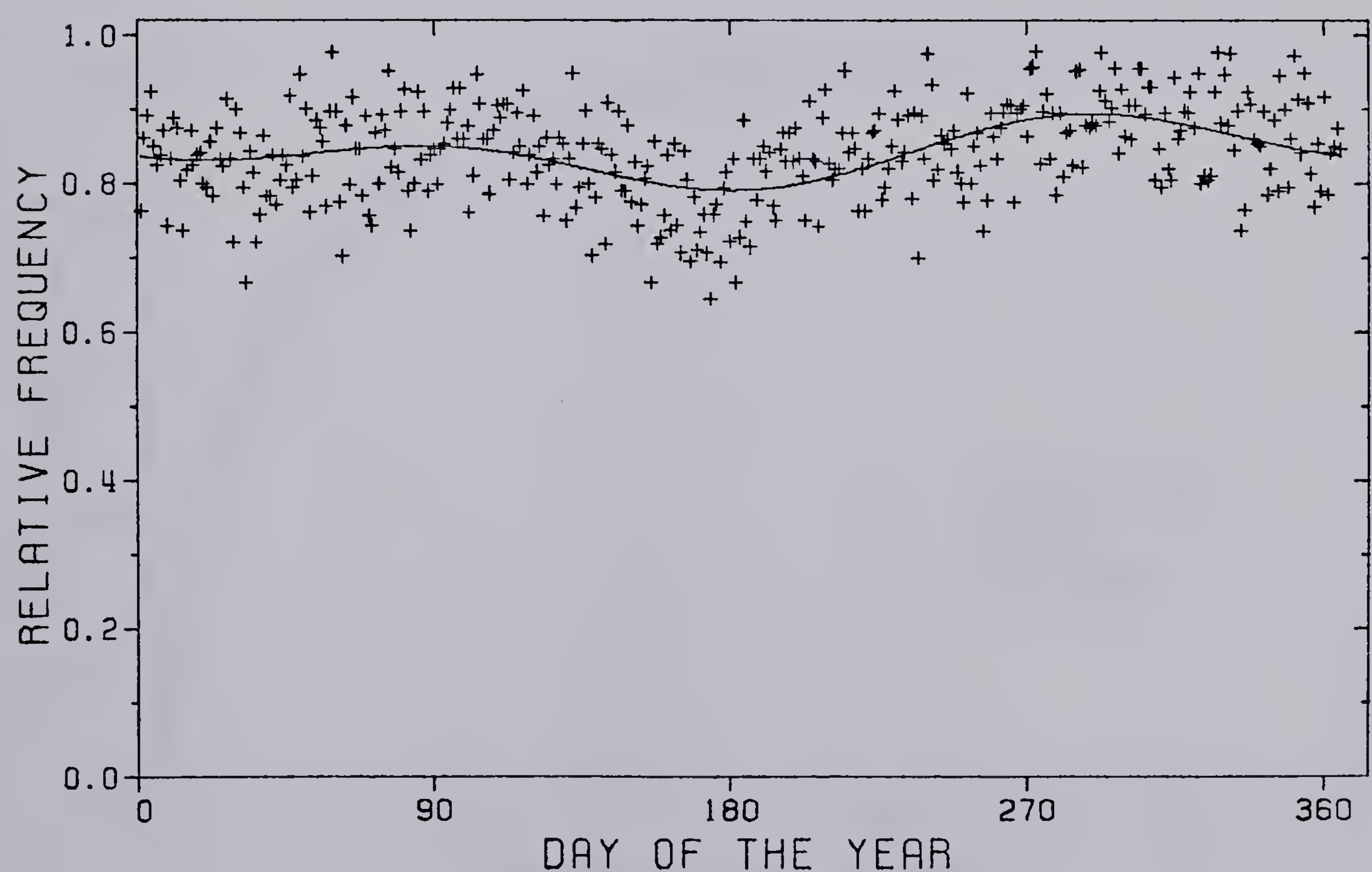


Figure 15. The Fourier series and raw estimates for P00 at Medicine Hat throughout the year.

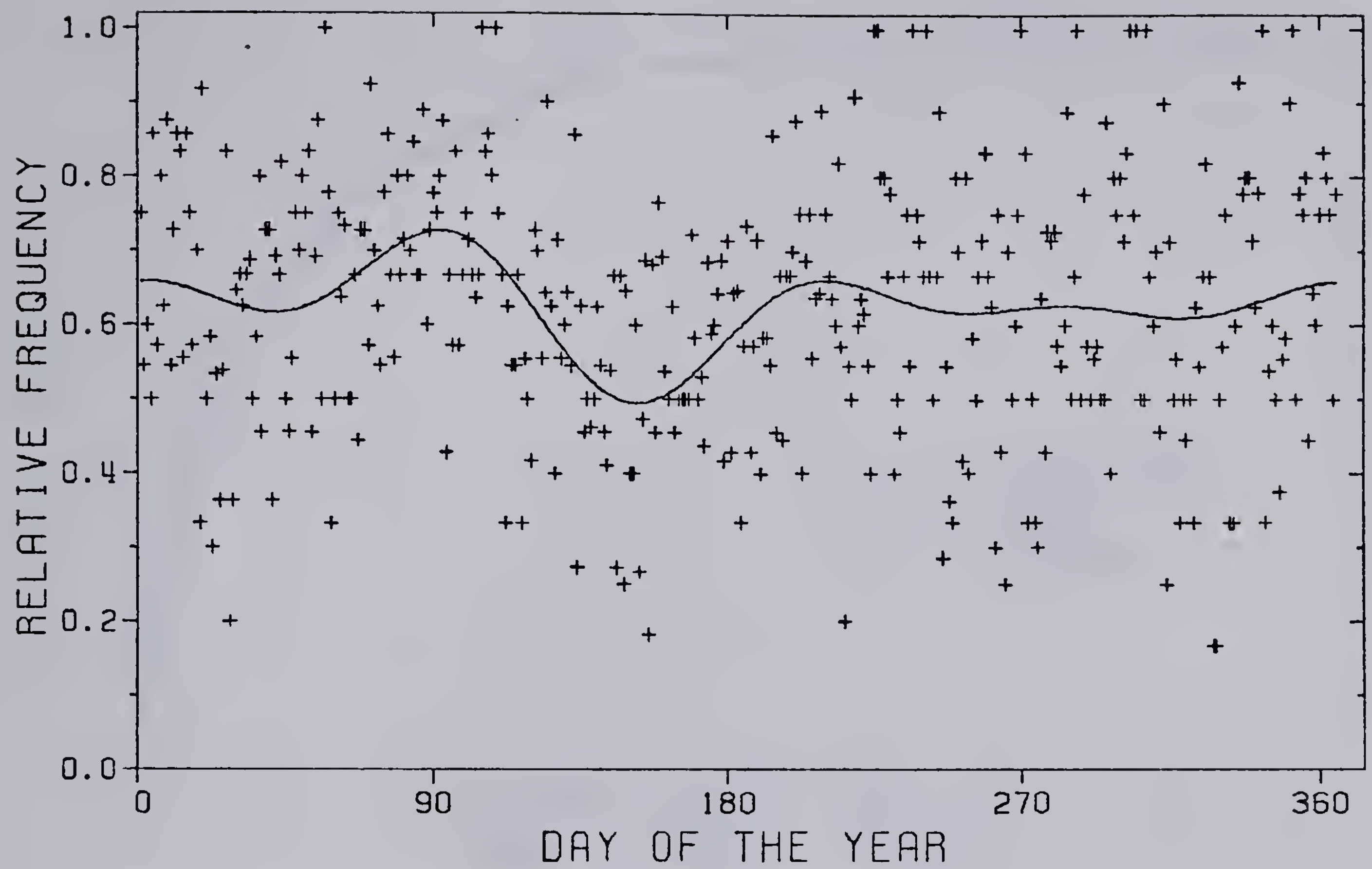


Figure 16. The Fourier series and raw estimates for P10 at Medicine Hat throughout the year.

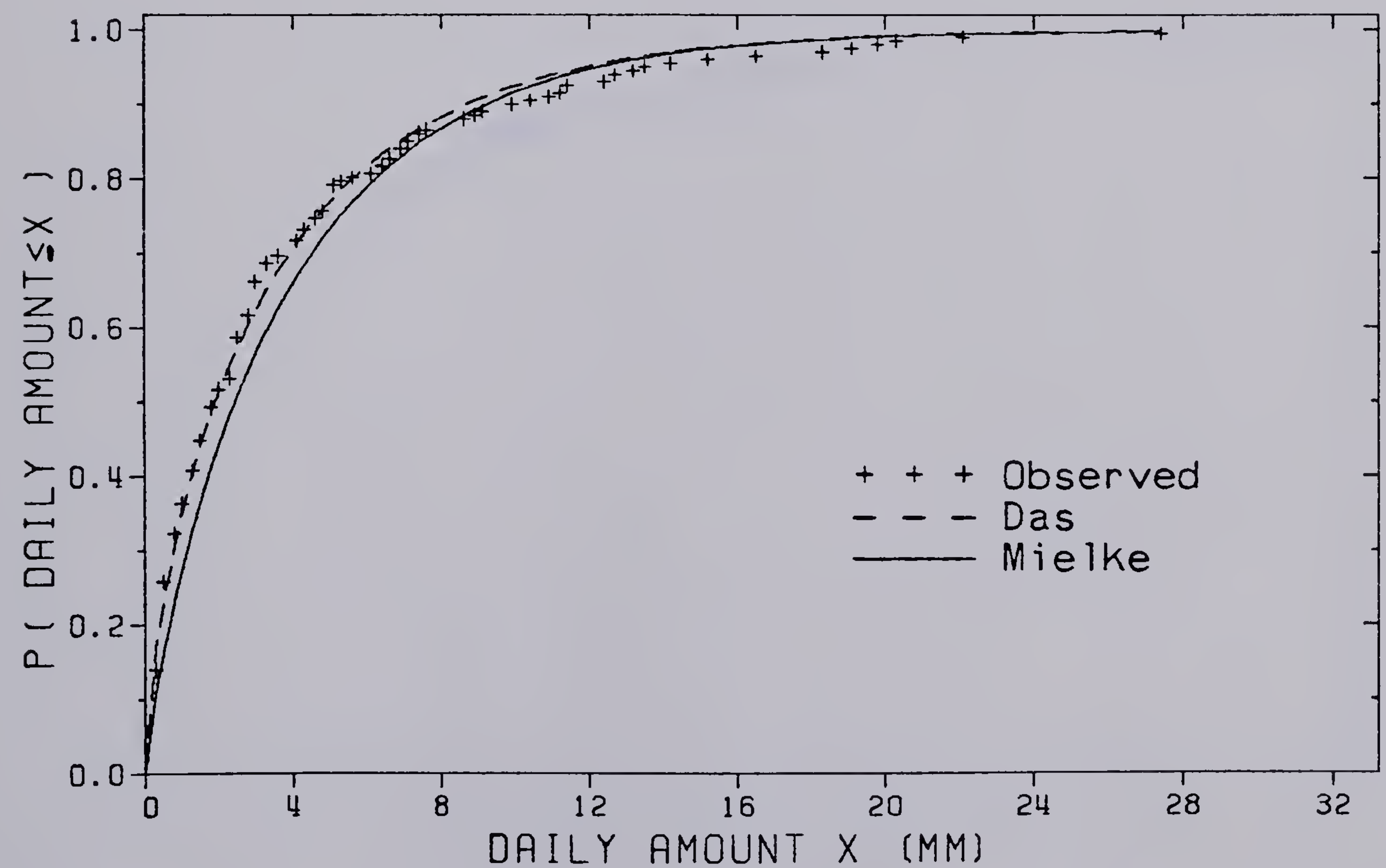


Figure 17. The observed and Gamma distributions for the daily amount of precipitation following a dry day during May at Beaverlodge.

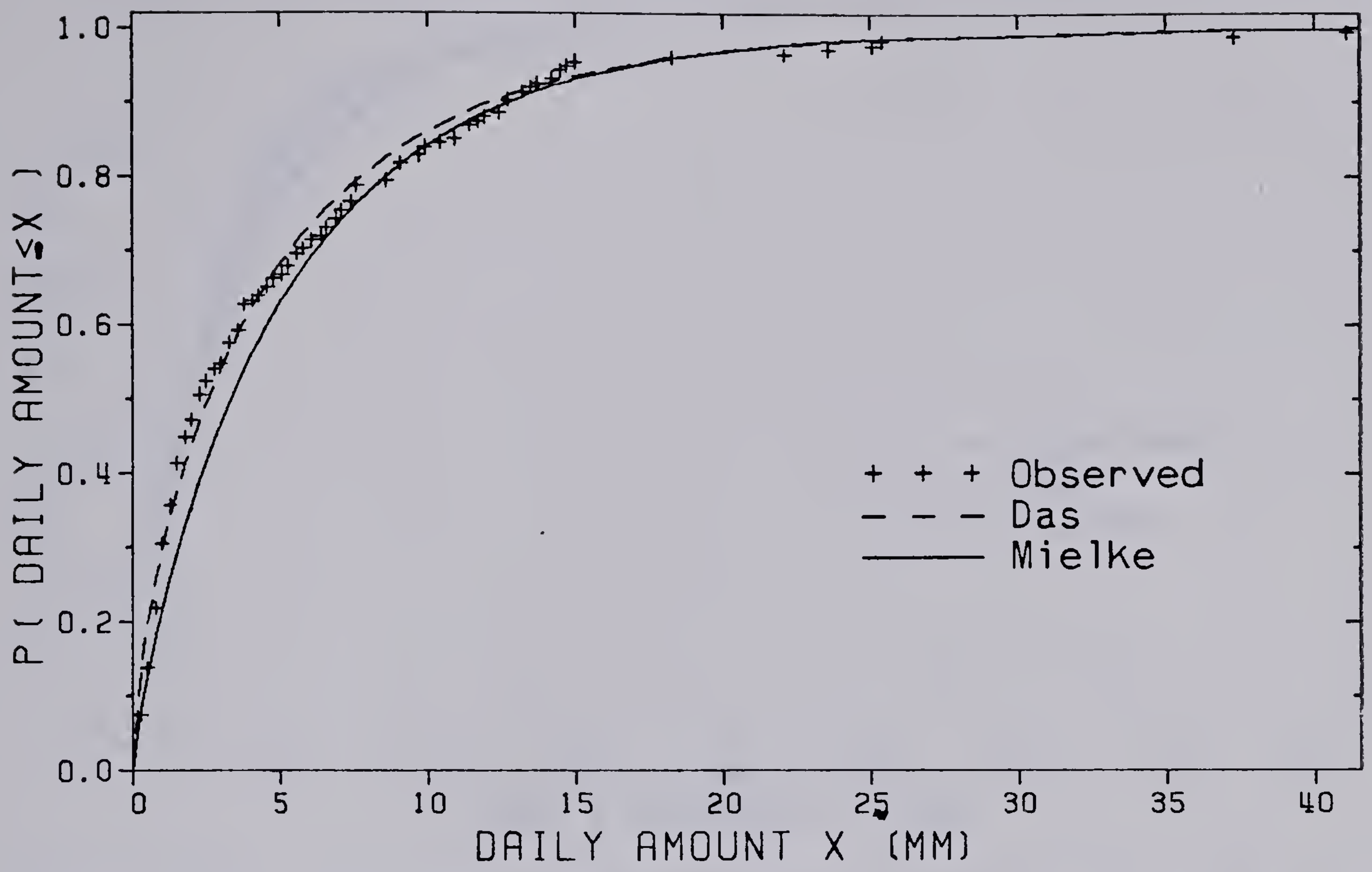


Figure 18. The observed and Gamma distributions for the daily amount of precipitation following a wet day during May at Beaverlodge.

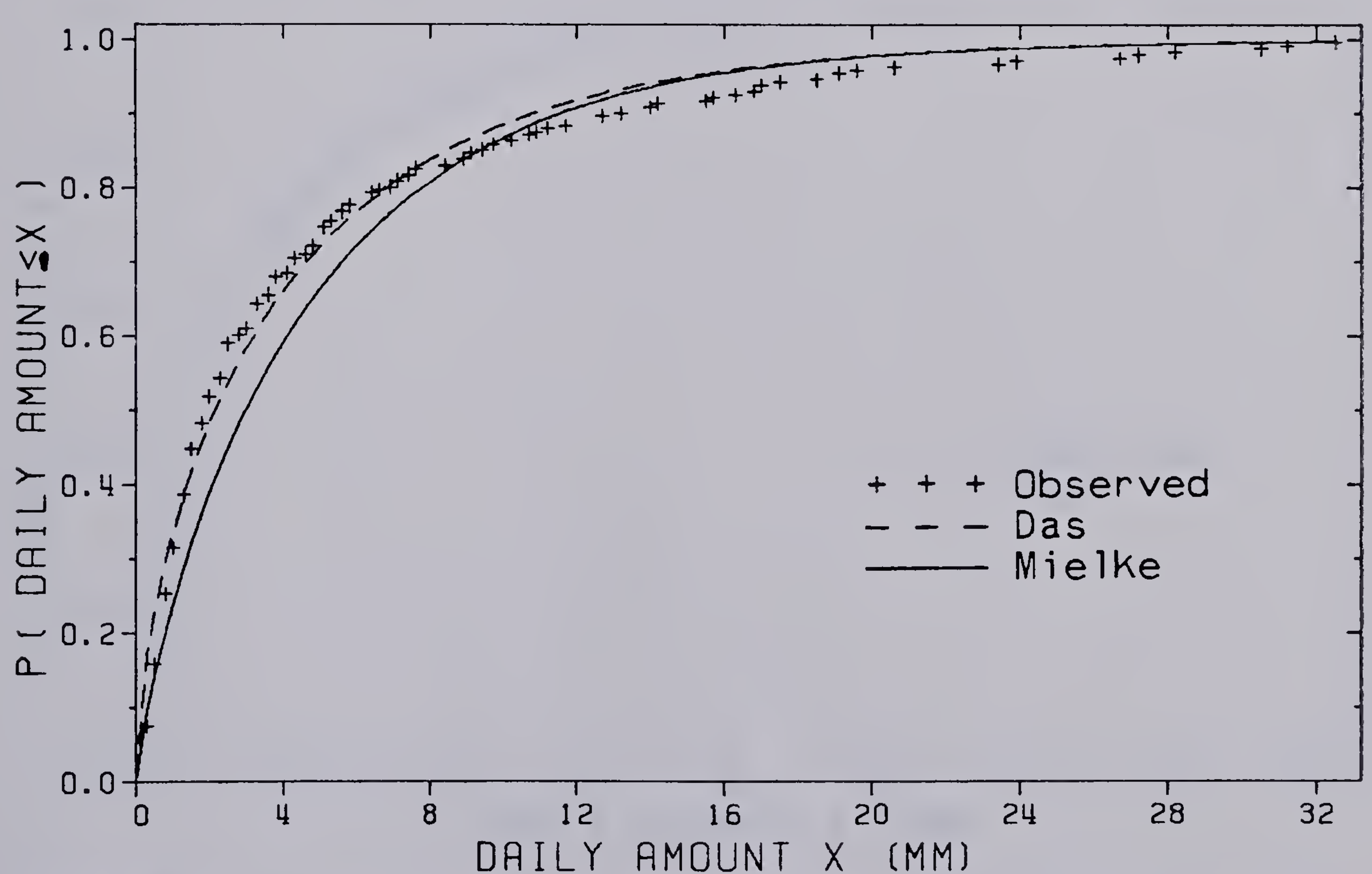


Figure 19. The observed and Gamma distributions for the daily amount of precipitation following a dry day during July at Beaverlodge.

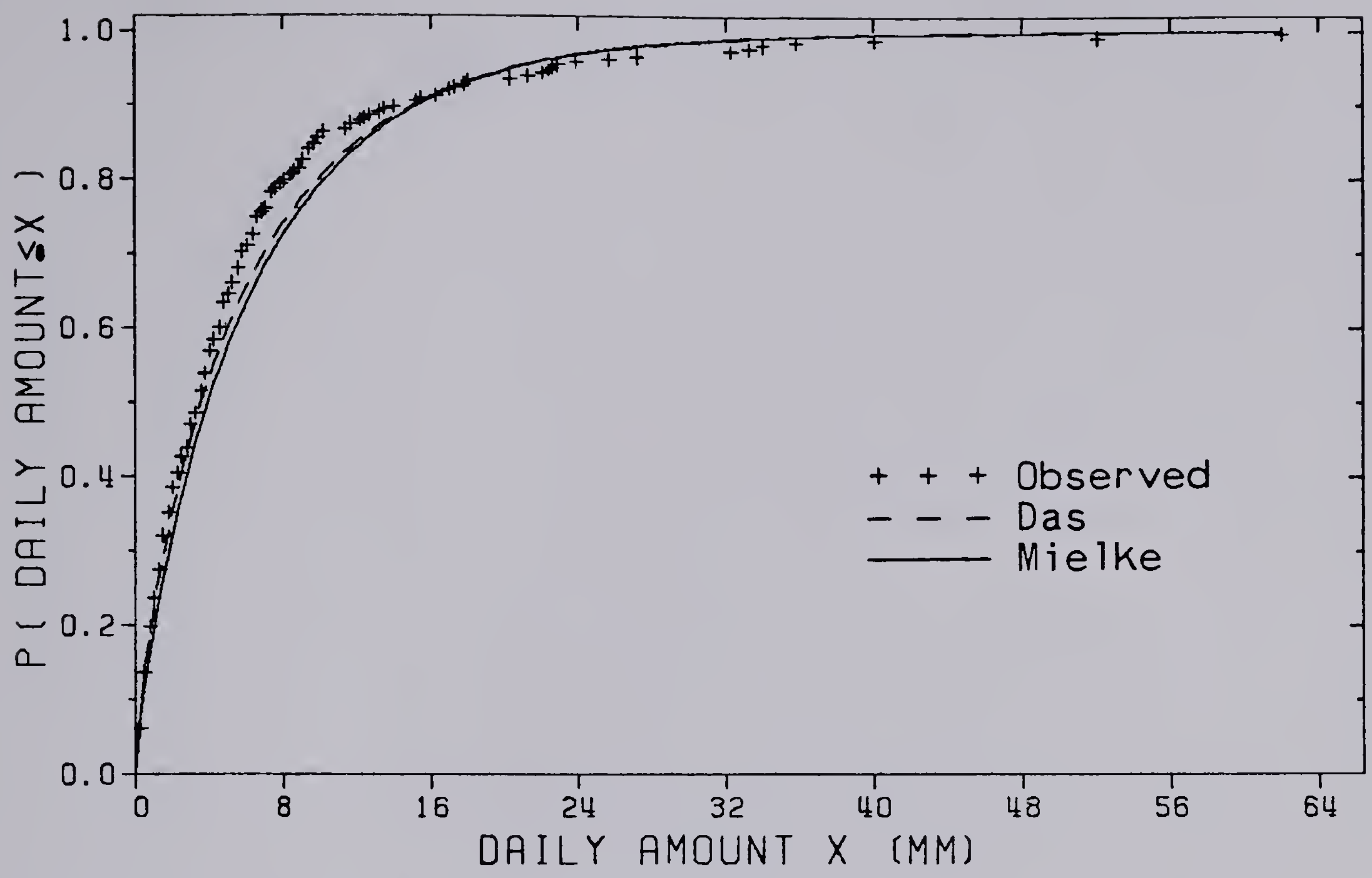


Figure 20. The observed and Gamma distributions for the daily amount of precipitation following a wet day during July at Beaverlodge.

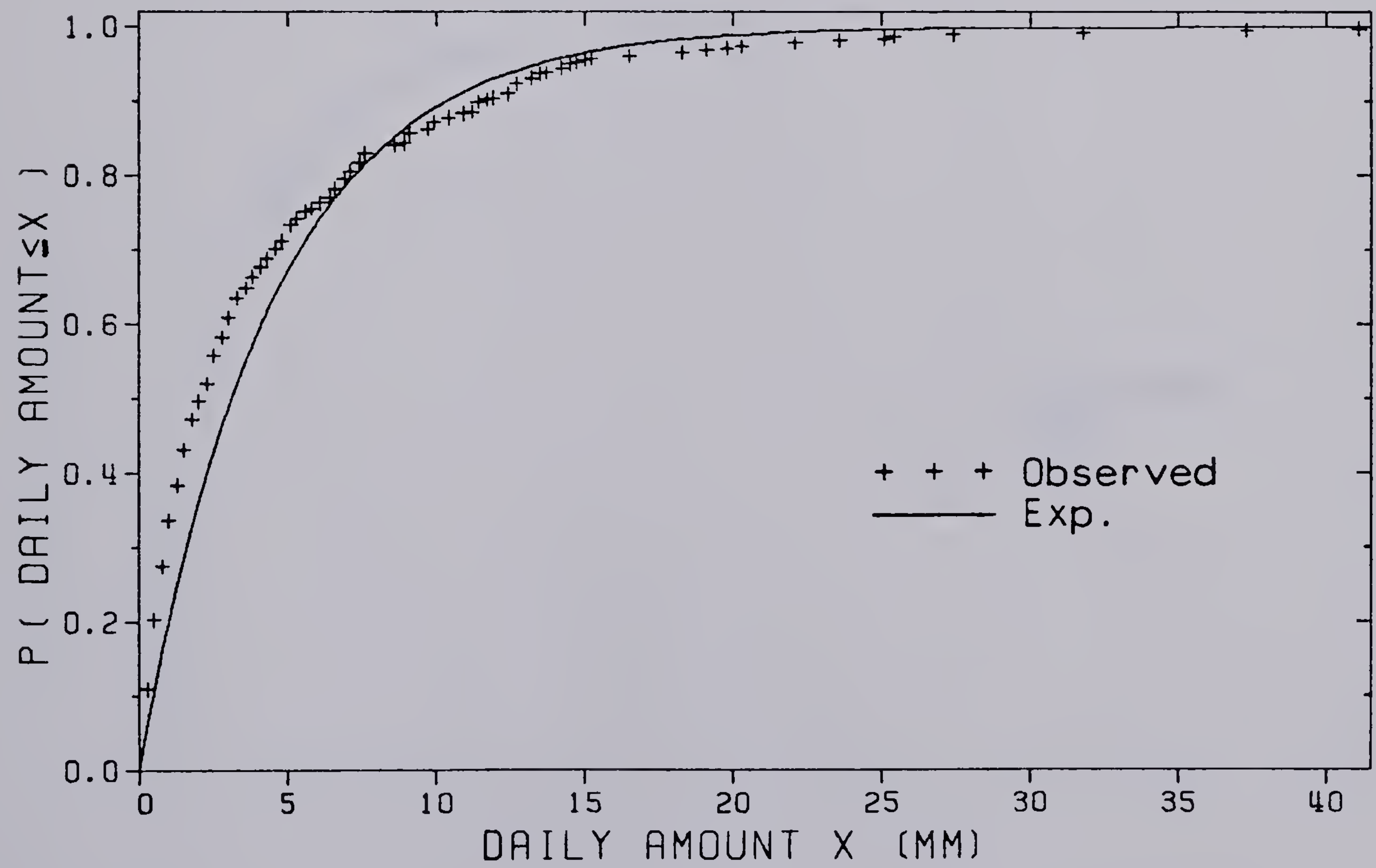


Figure 21. The observed and exponential distributions for the daily amount of precipitation during May at Beaverlodge.

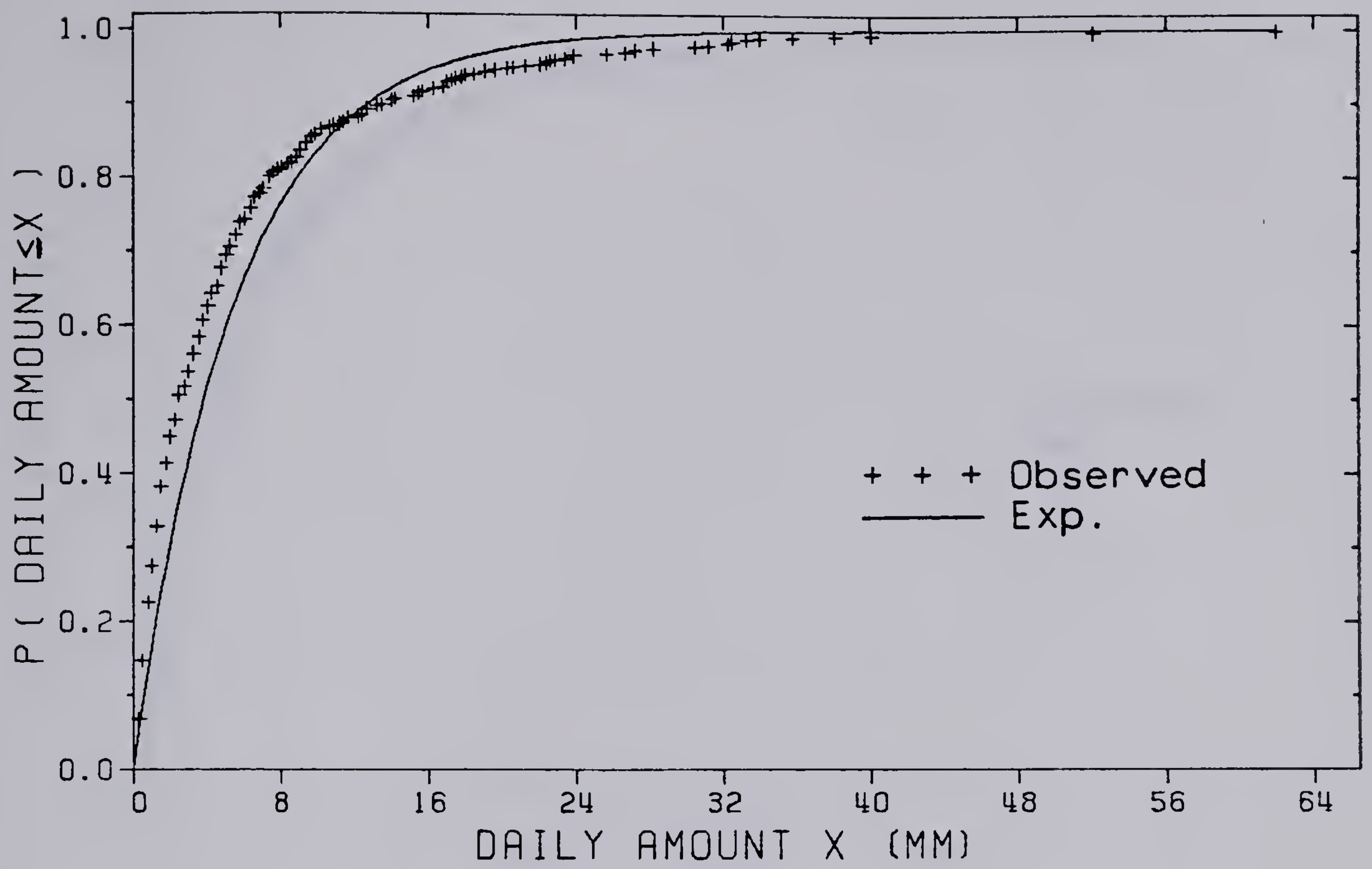


Figure 22. The observed and exponential distributions for the daily amount of precipitation during July at Beaverlodge.

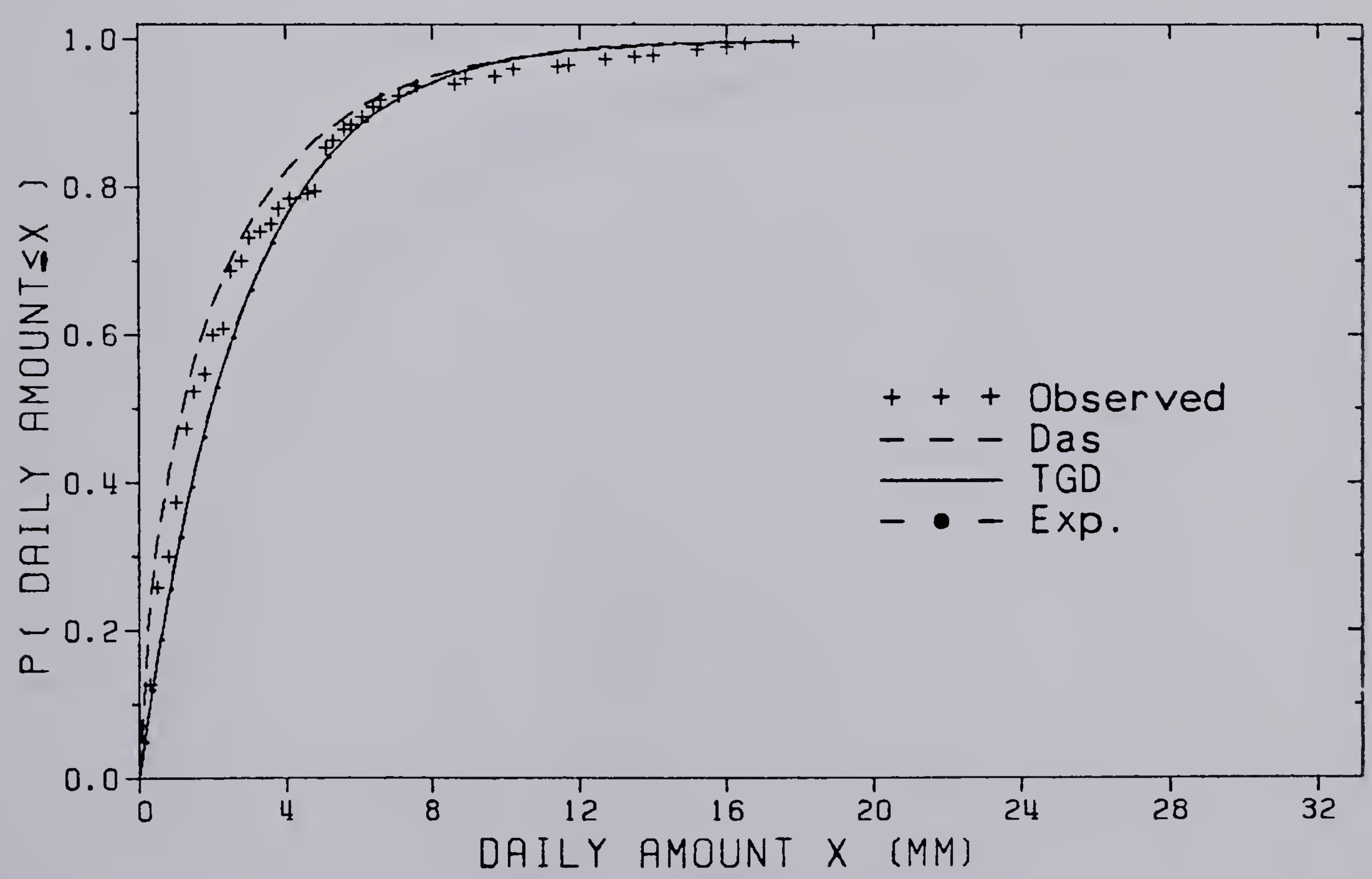


Figure 23. The observed and theoretical distributions for the daily amount of precipitation during January at Edmonton.

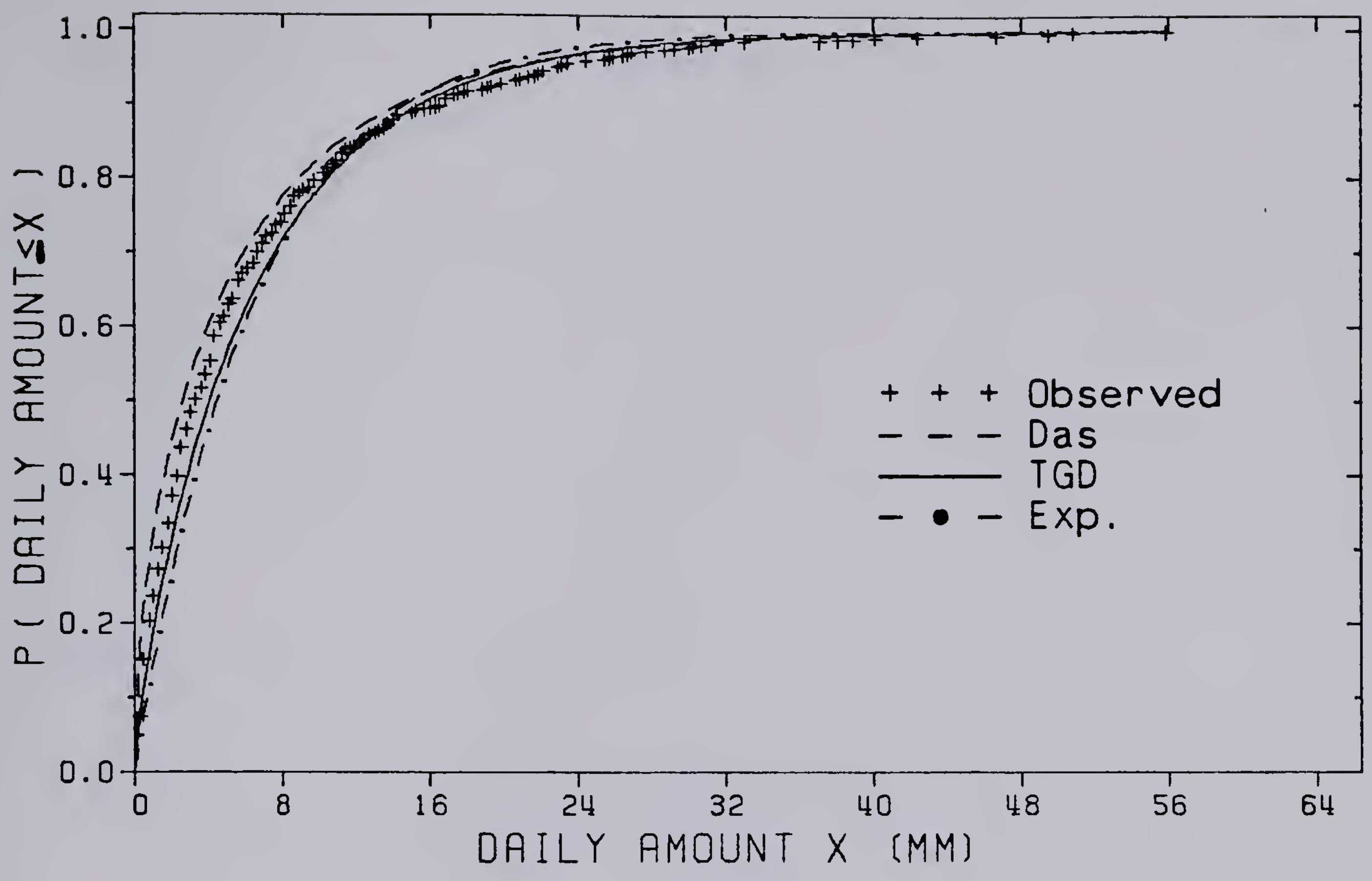


Figure 24. The observed and theoretical distributions for the daily amount of precipitation during June at Edmonton.

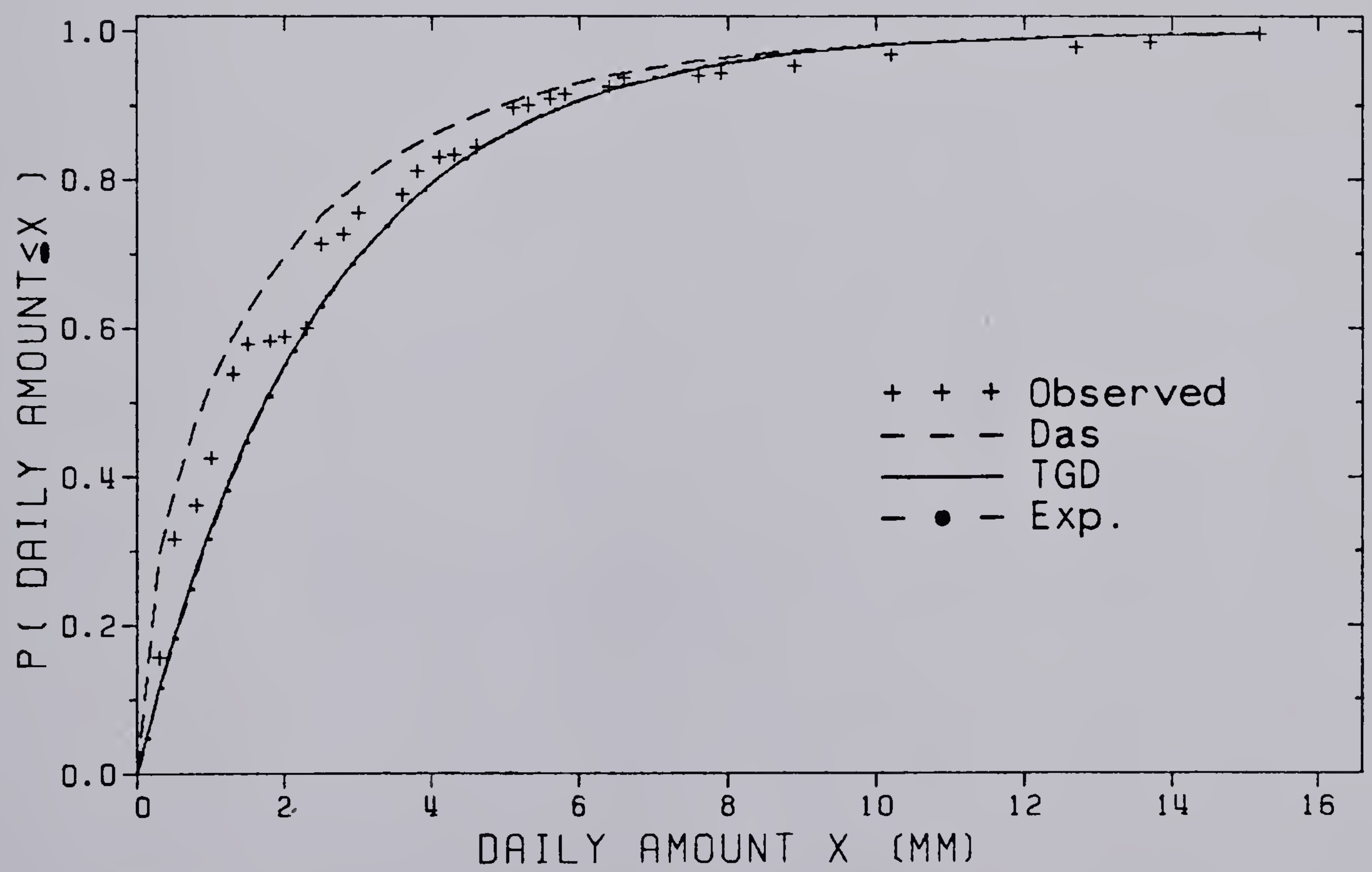


Figure 25. The observed and theoretical distributions for the daily amount of precipitation during March at Medicine Hat.

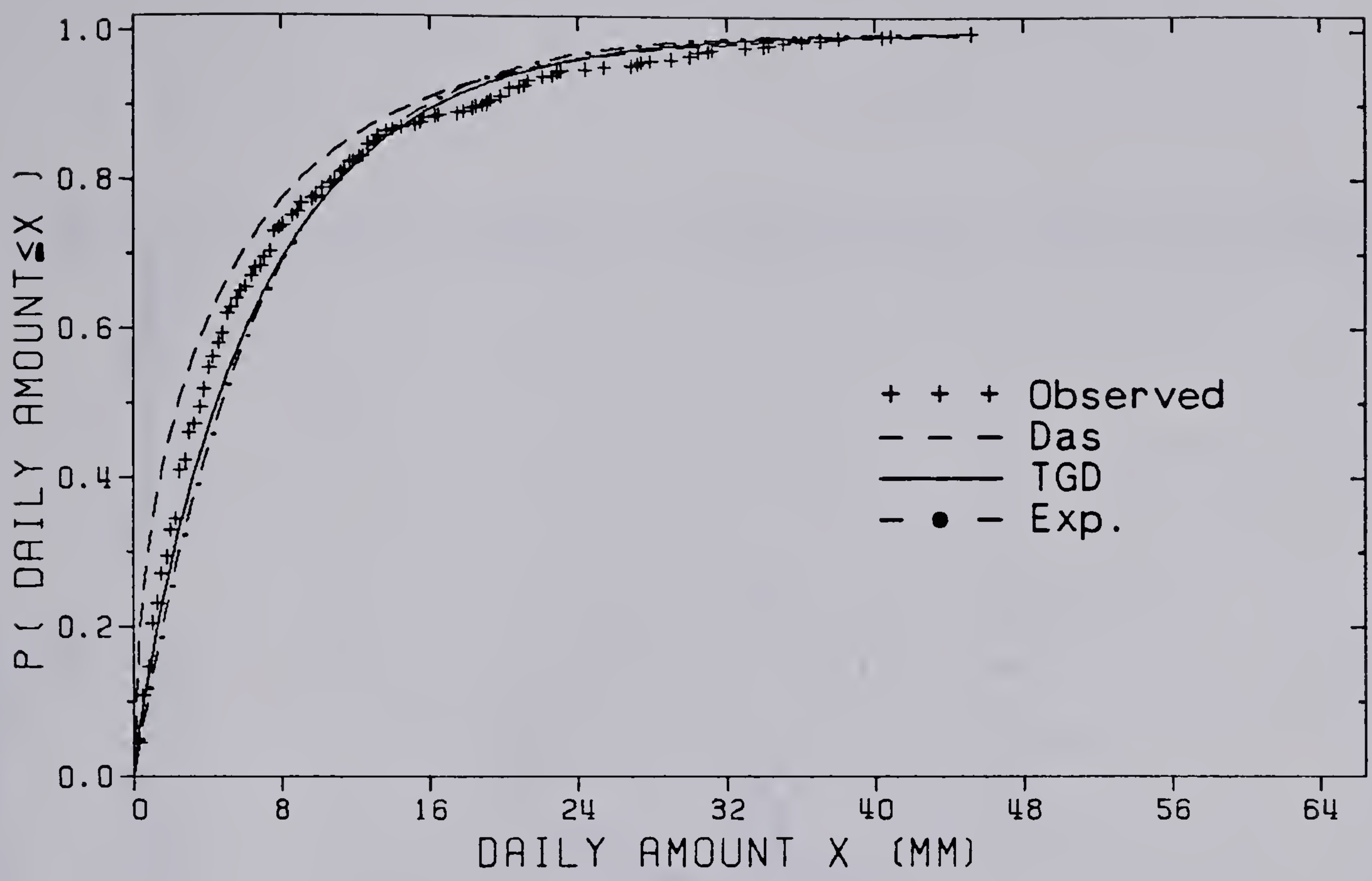


Figure 26. The observed and theoretical distributions for the daily amount of precipitation during June at Medicine Hat.

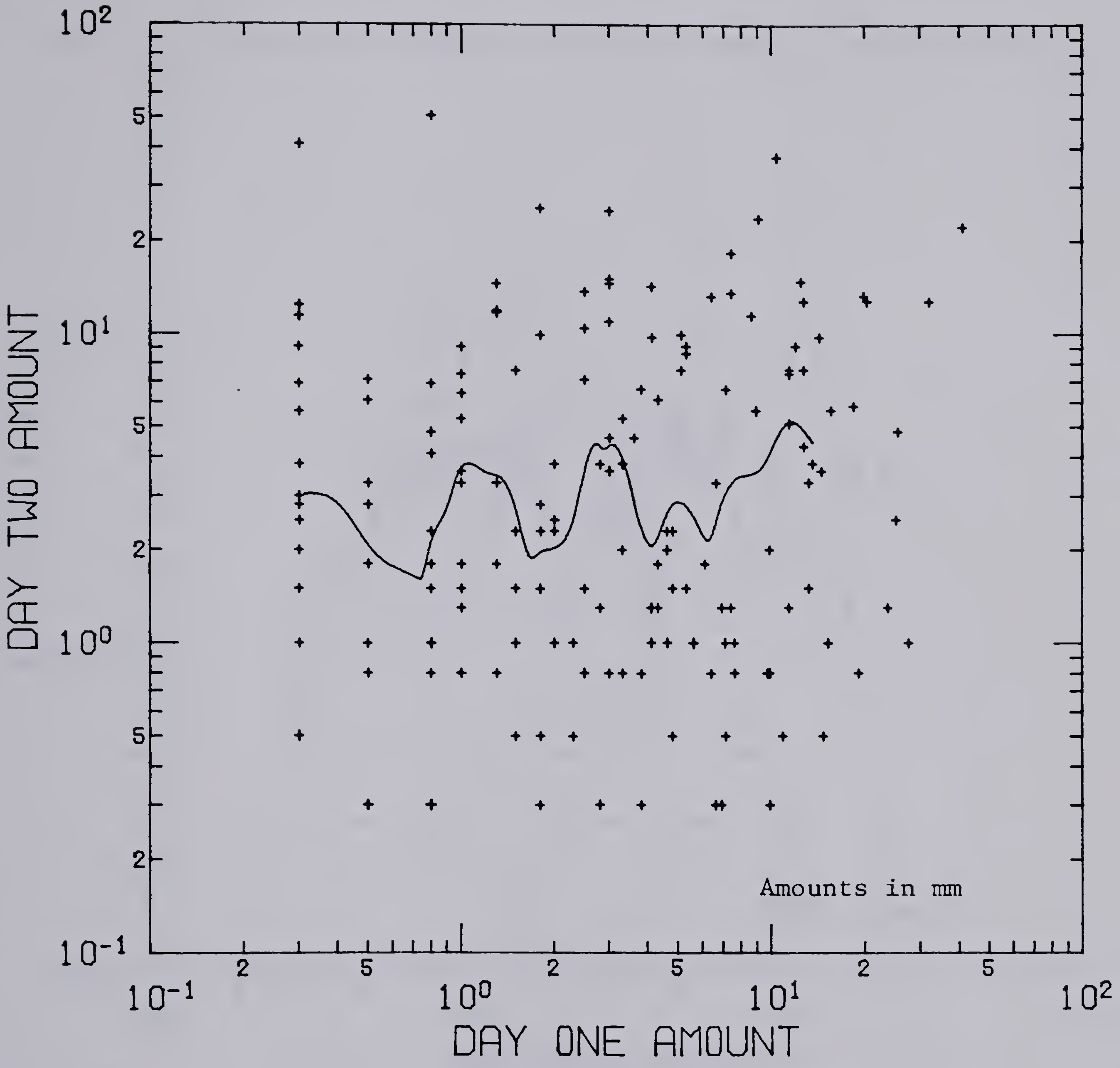


Figure 27. The first daily amount of precipitation versus the second for consecutive wet days in May at Beaverlodge.

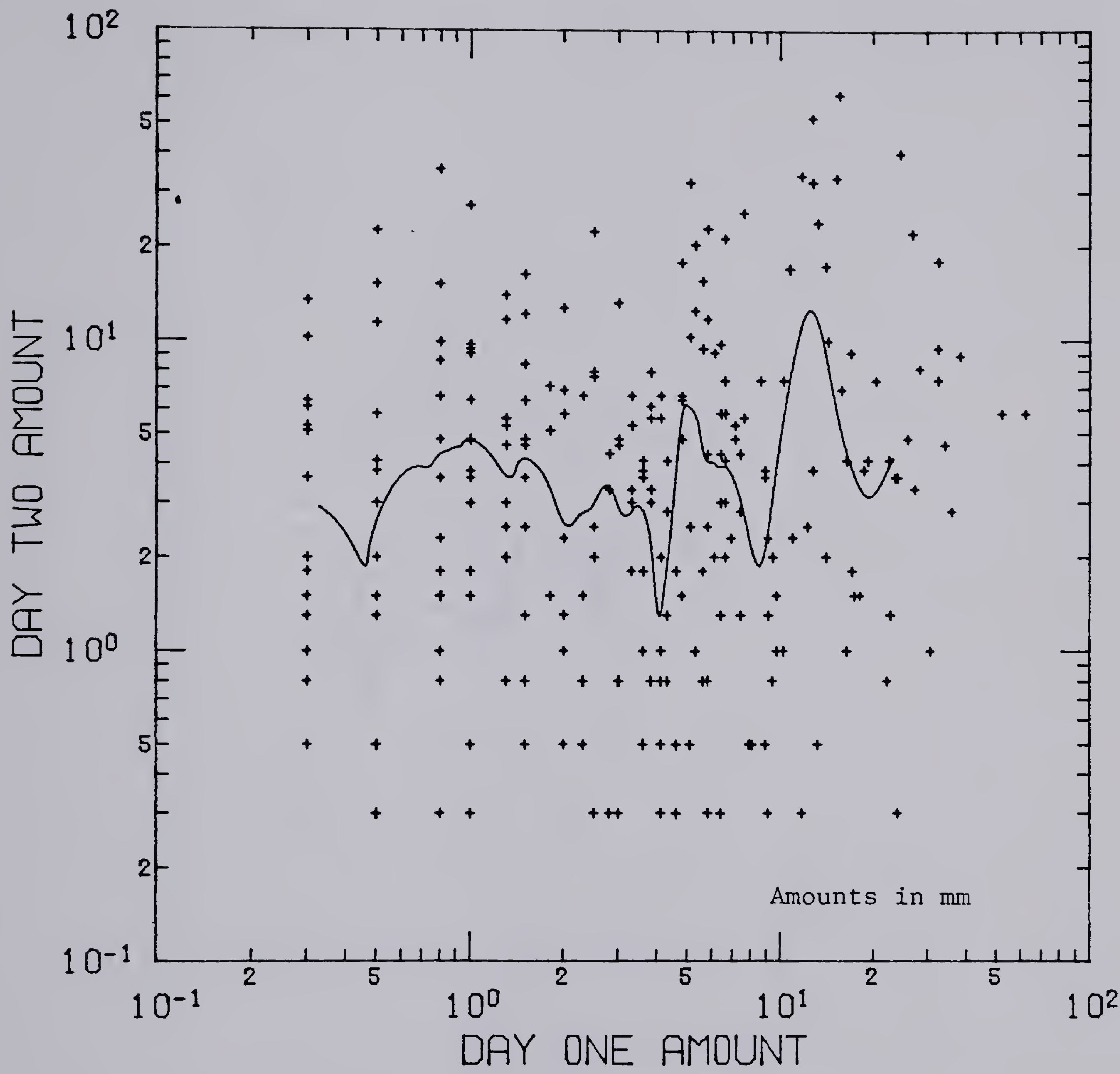


Figure 28. The first daily amount of precipitation versus the second for consecutive wet days in July at Beaverlodge.

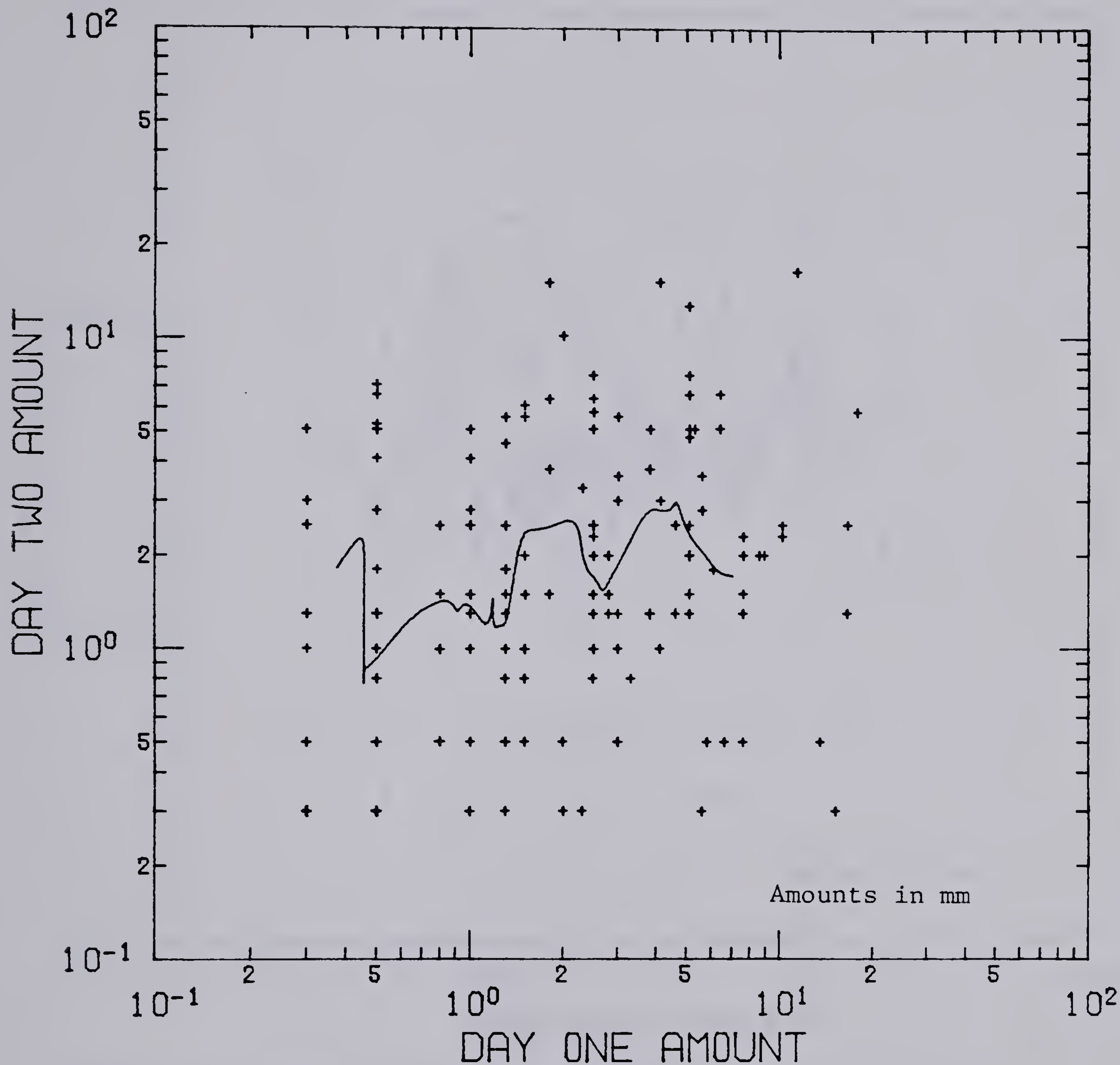


Figure 29. The first daily amount of precipitation versus the second for consecutive wet days in January at Edmonton.

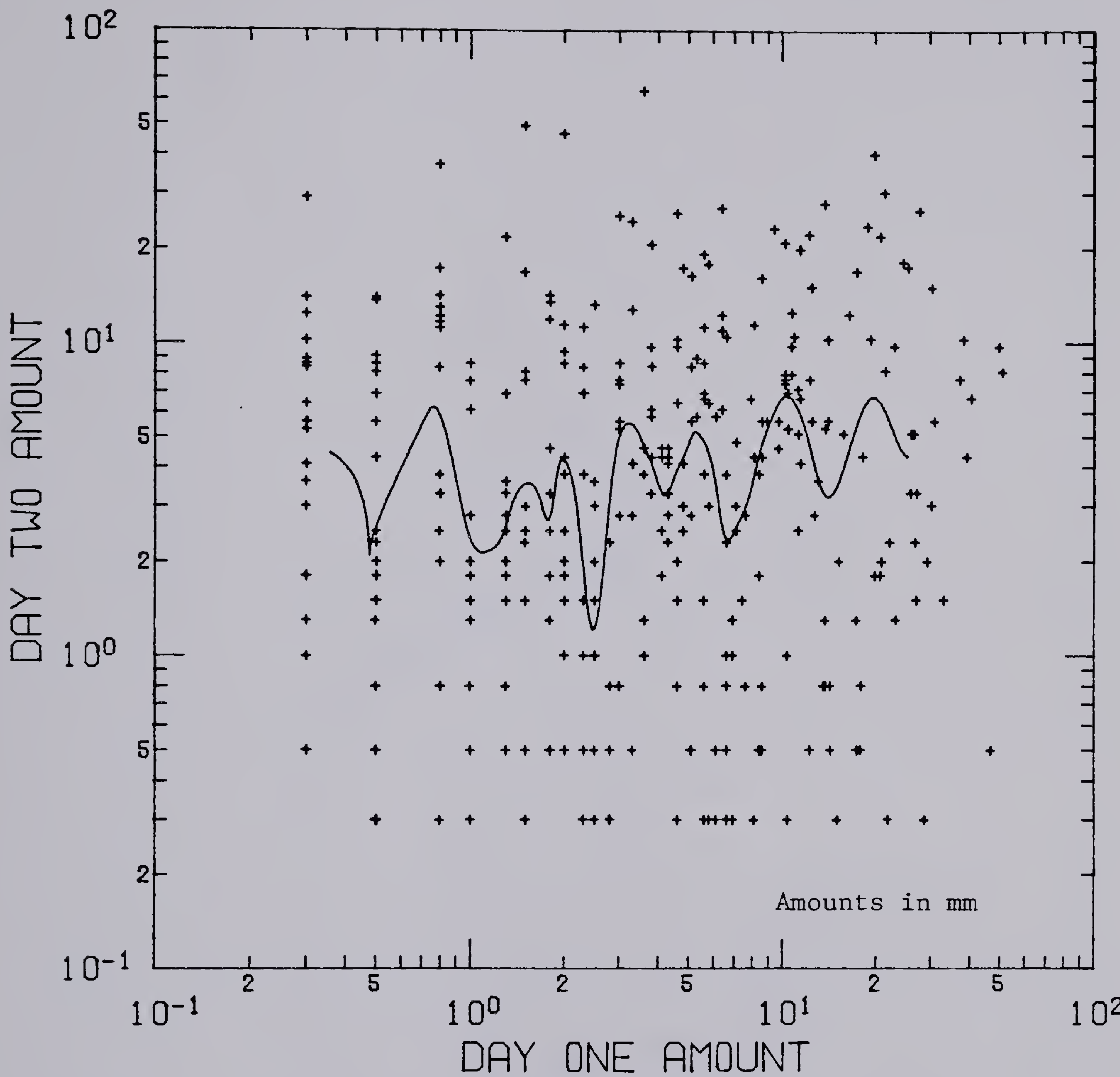


Figure 30. The first daily amount of precipitation versus the second for consecutive wet days in June at Edmonton.

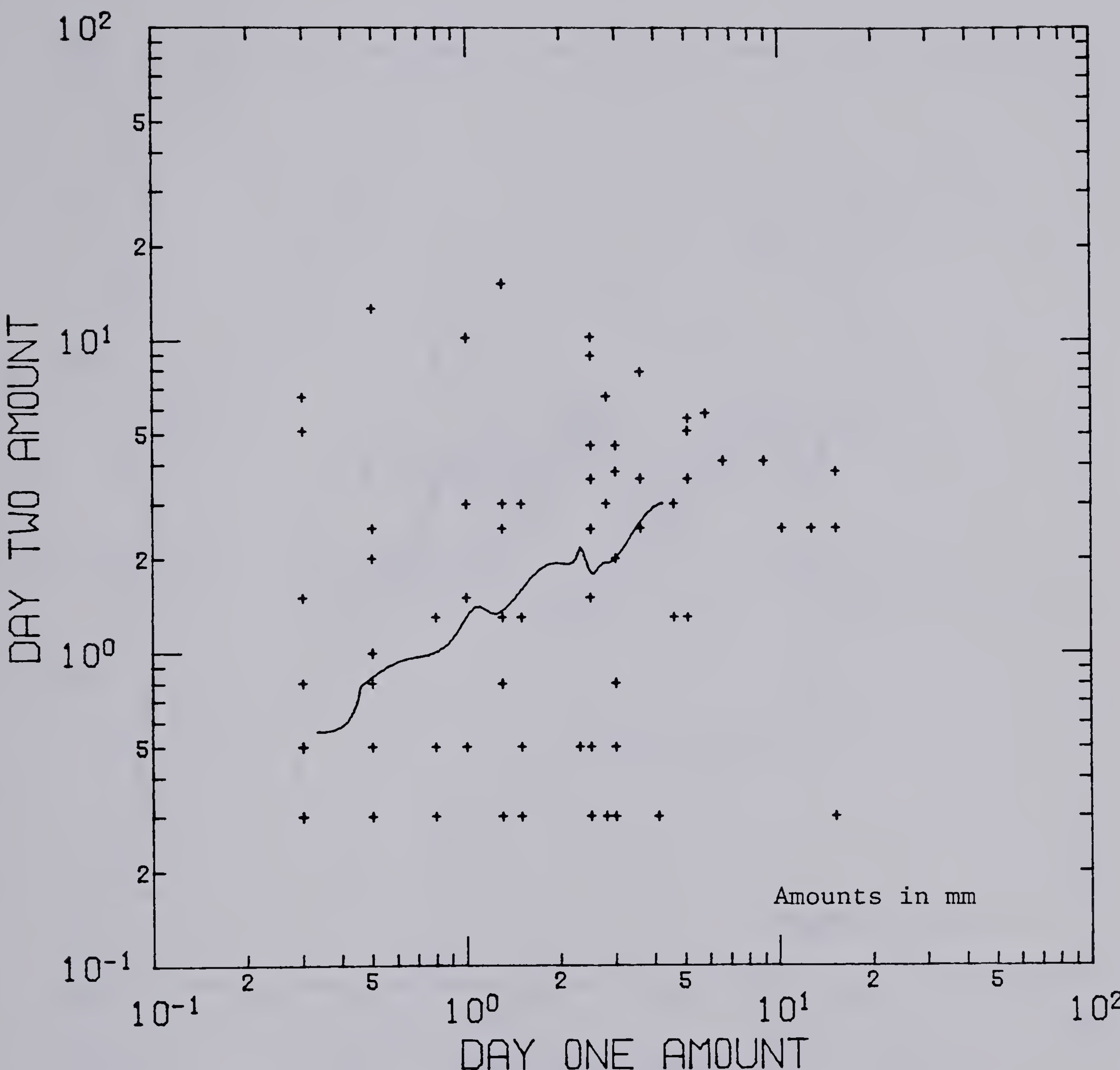


Figure 31. The first daily amount of precipitation versus the second for consecutive wet days in March at Medicine Hat.

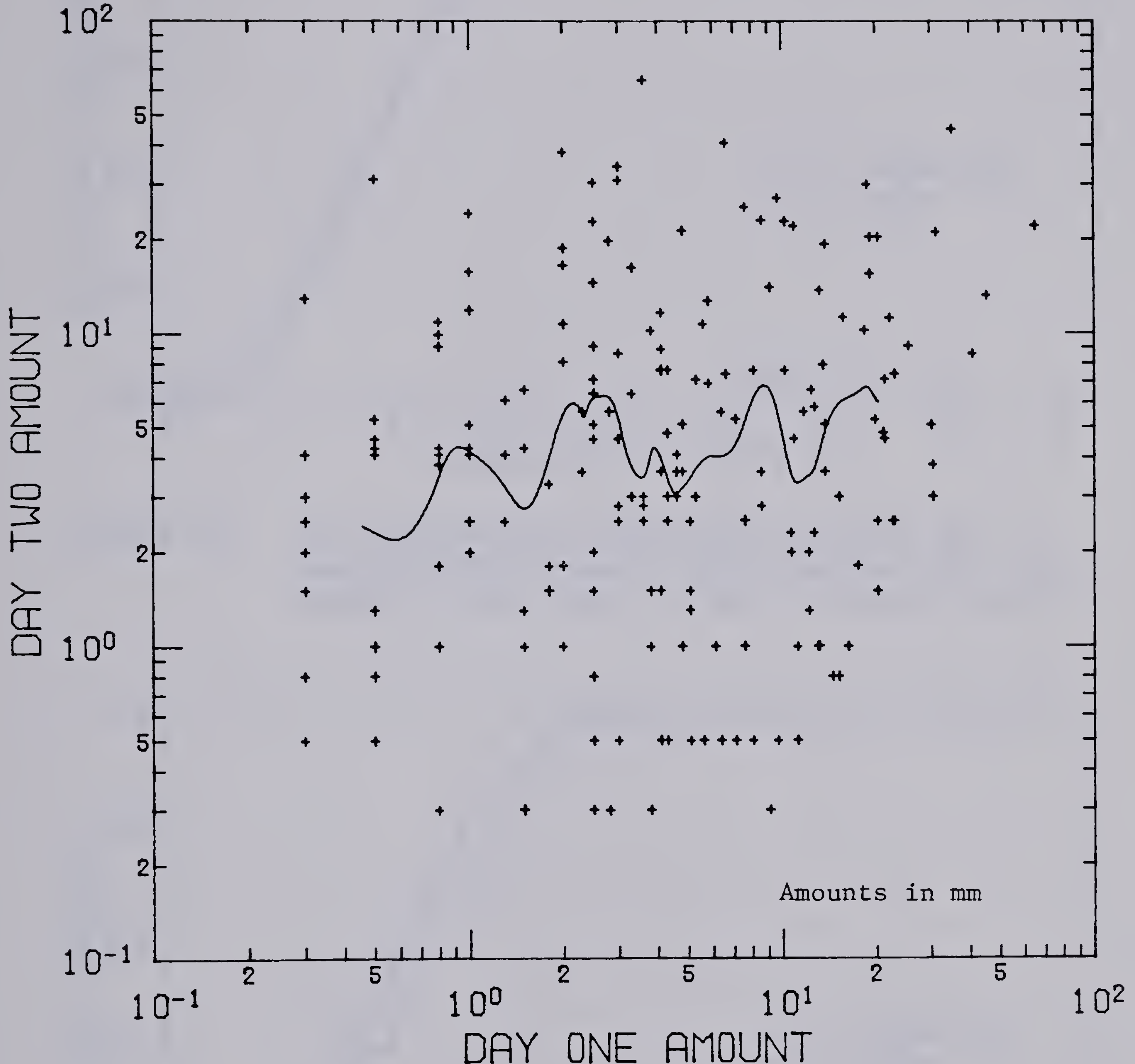


Figure 32. The first daily amount of precipitation versus the second for consecutive wet days in June at Medicine Hat.

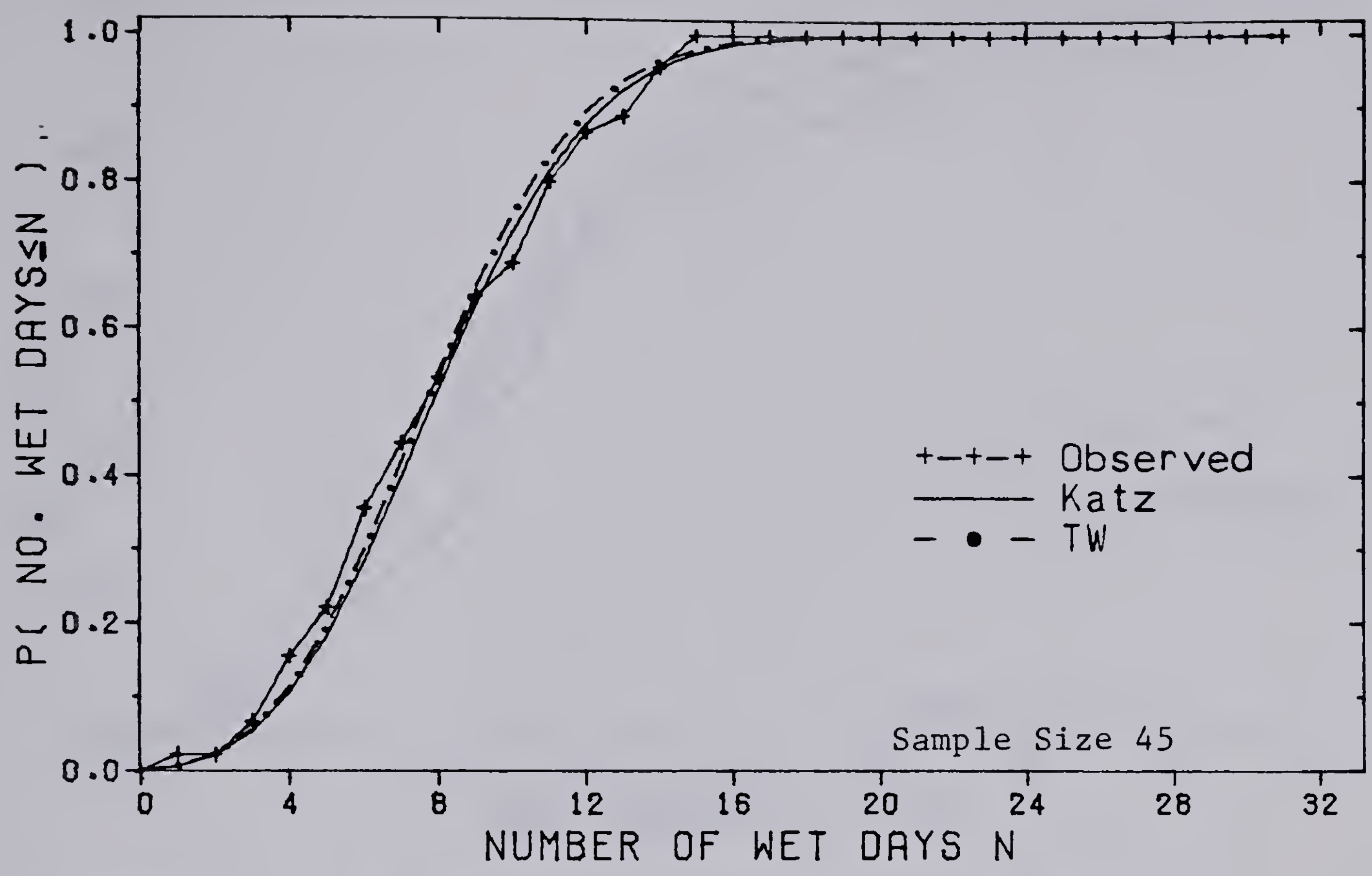


Figure 33. The theoretical distributions and the observed development distribution for the number of wet days in May at Beaverlodge.

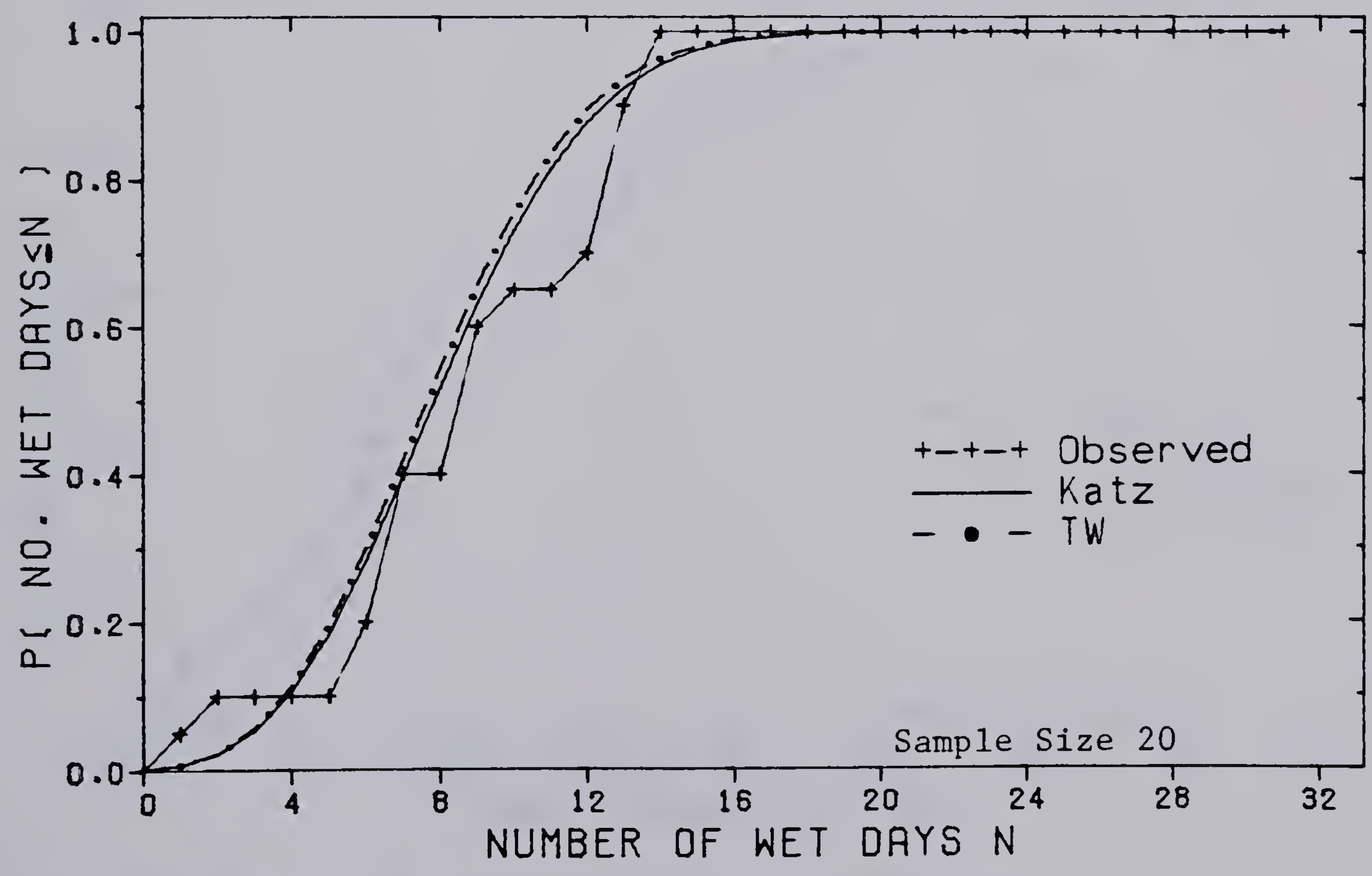


Figure 34. The theoretical distributions and the observed independent distribution for the number of wet days in May at Beaverlodge.

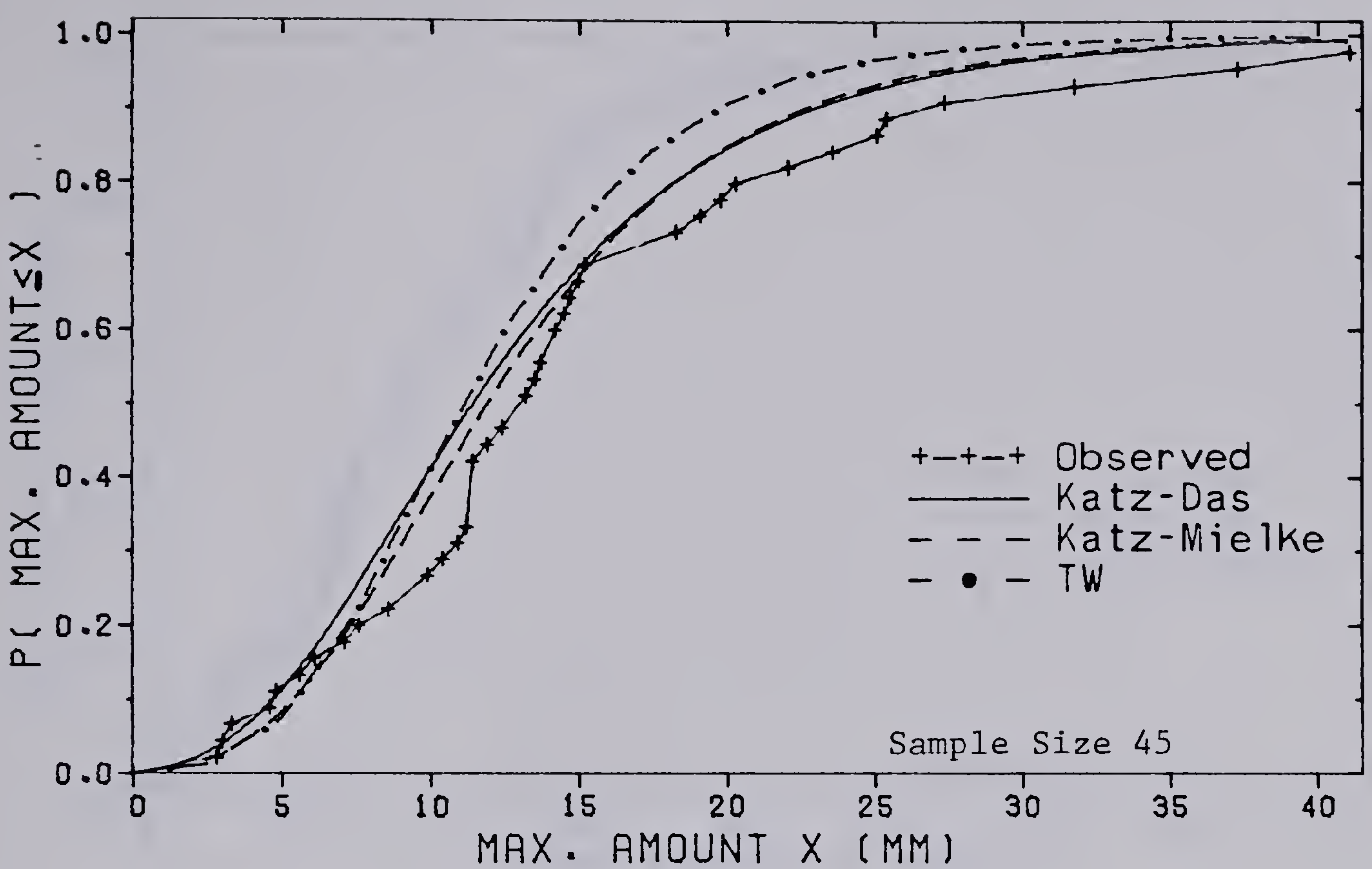


Figure 35. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in May at Beaverlodge.

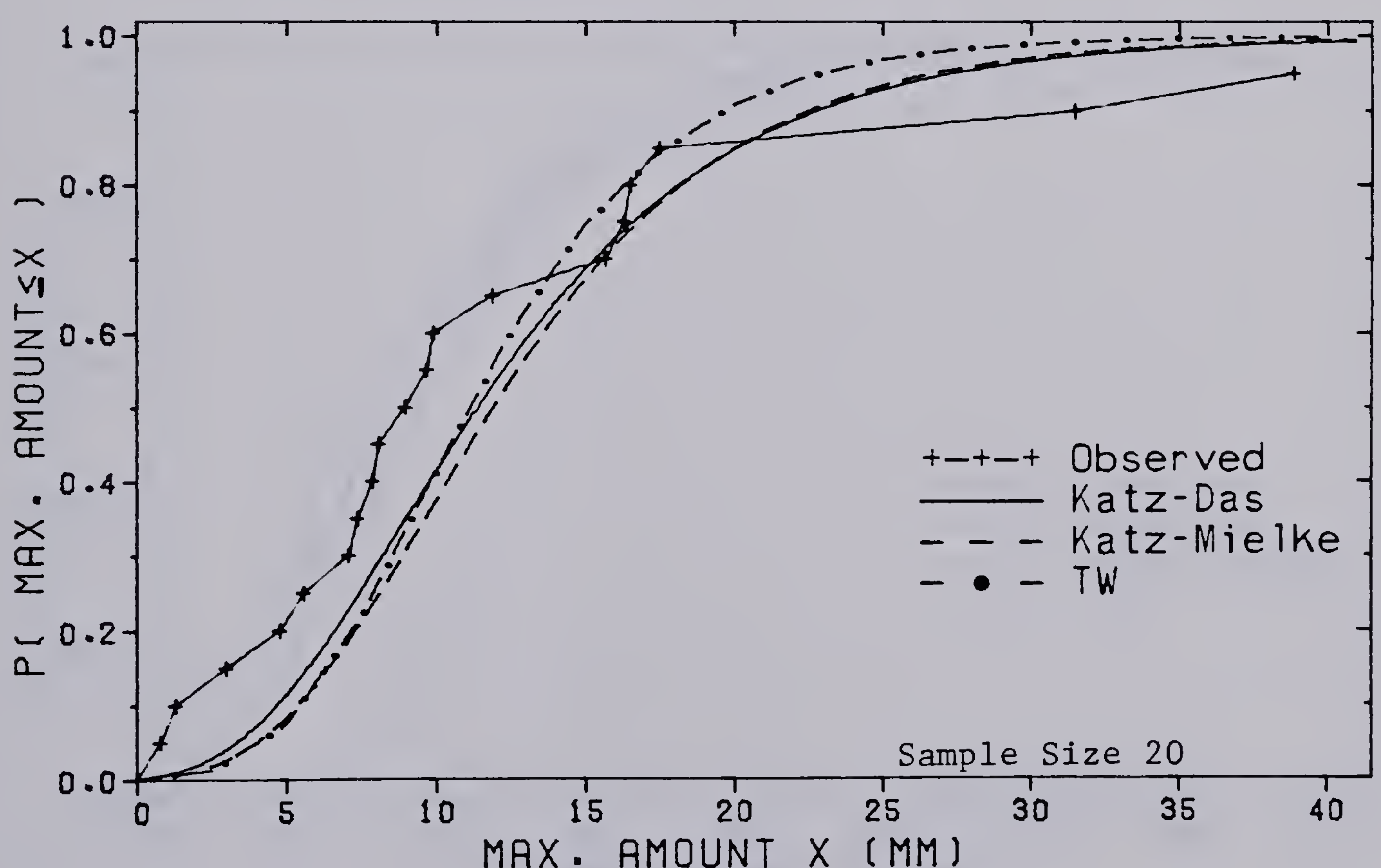


Figure 36. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in May at Beaverlodge.

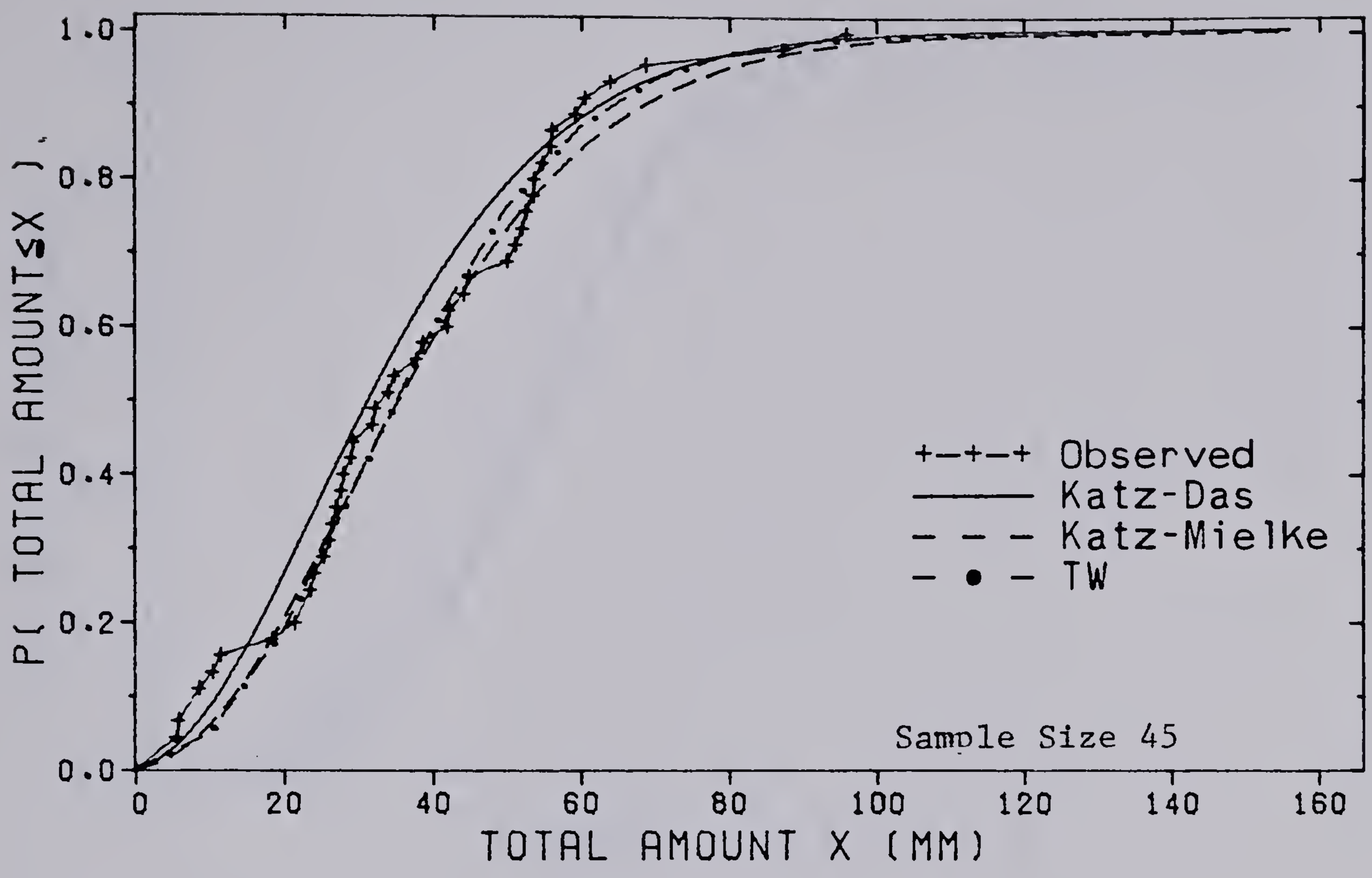


Figure 37. The theoretical distributions and the observed development distribution for the total amount of precipitation in May at Beaverlodge.

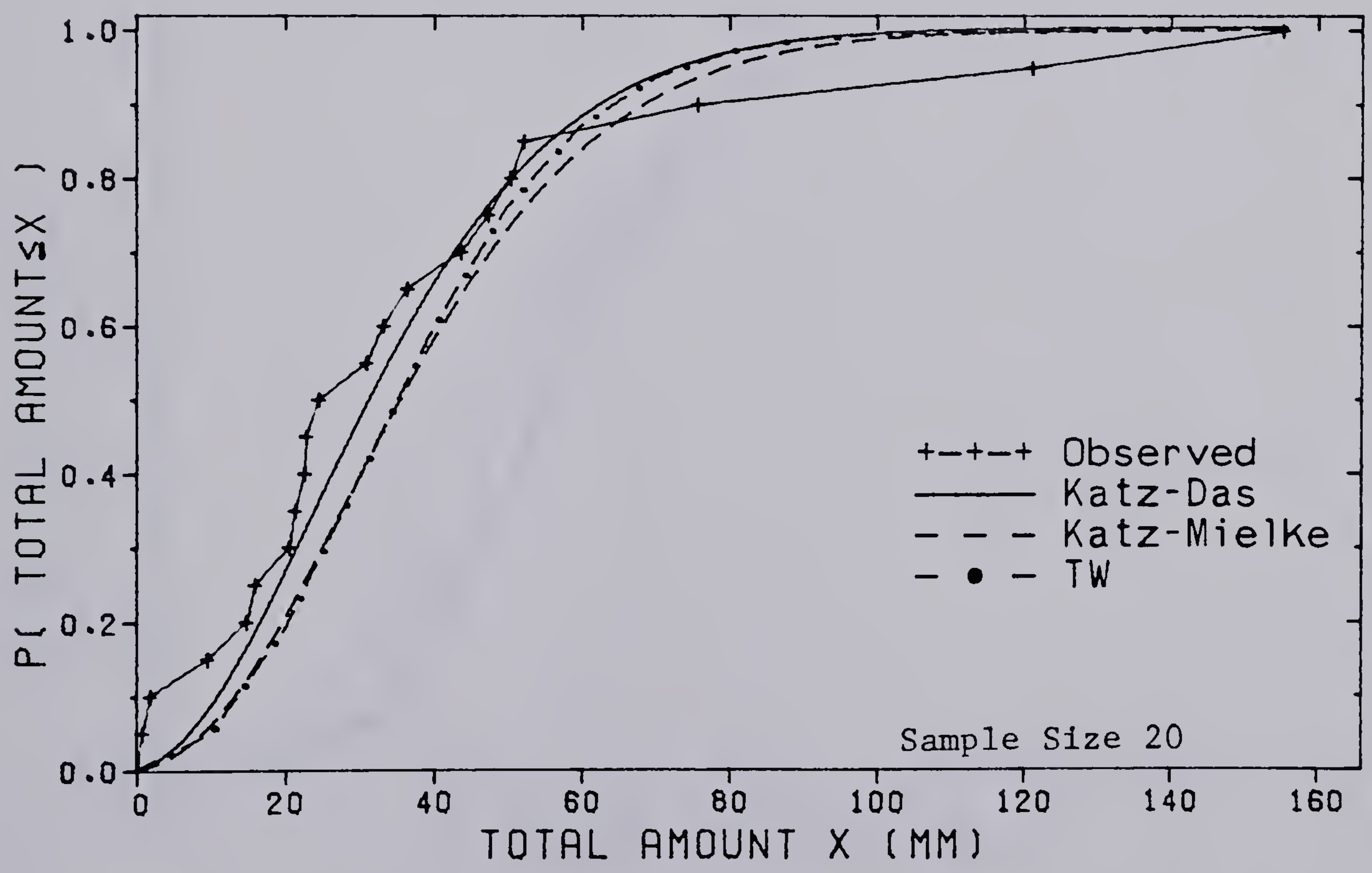


Figure 38. The theoretical distributions and the observed independent distribution for the total amount of precipitation in May at Beaverlodge.

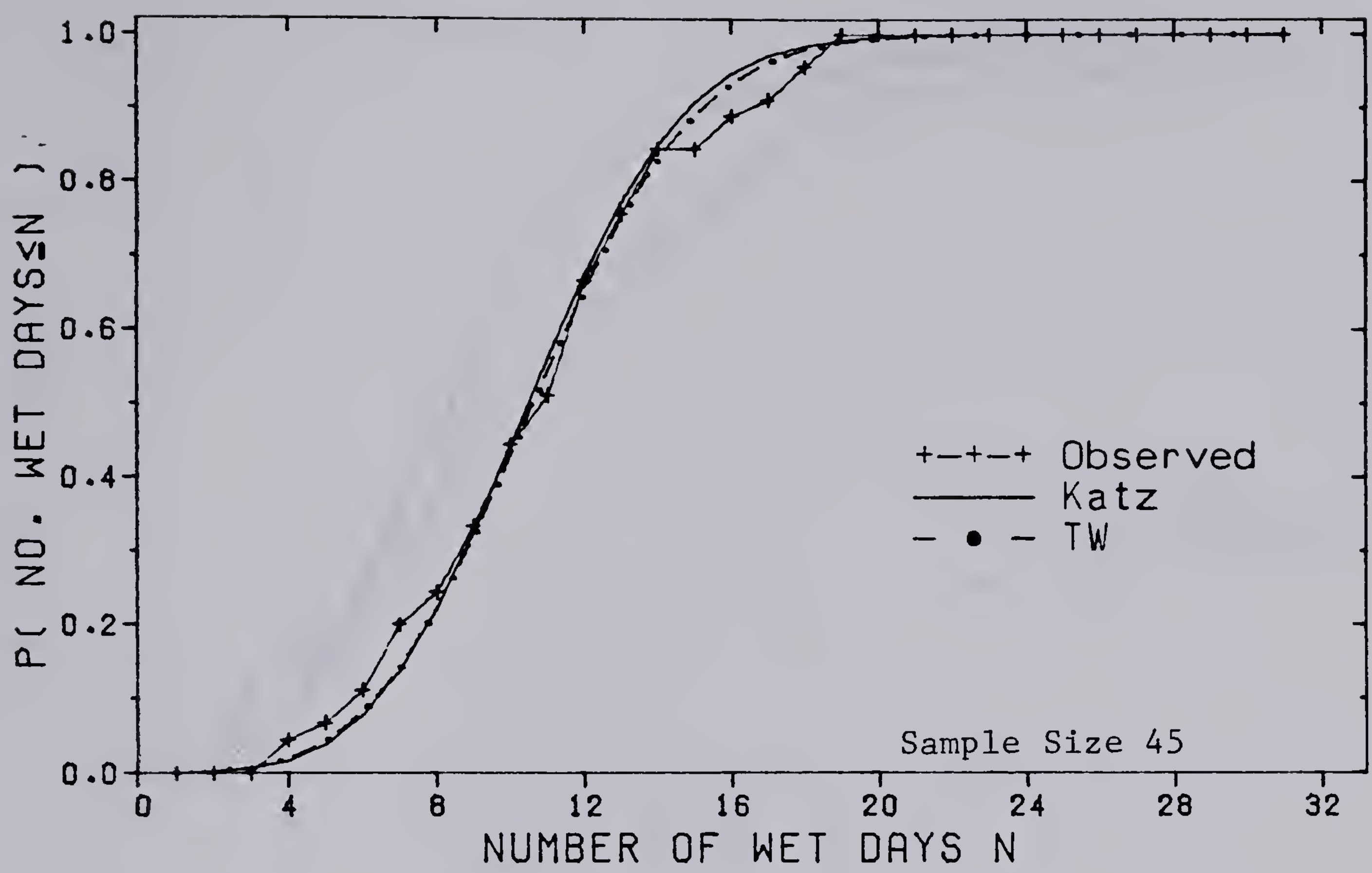


Figure 39. The theoretical distributions and the observed development distribution for the number of wet days in July at Beaverlodge.

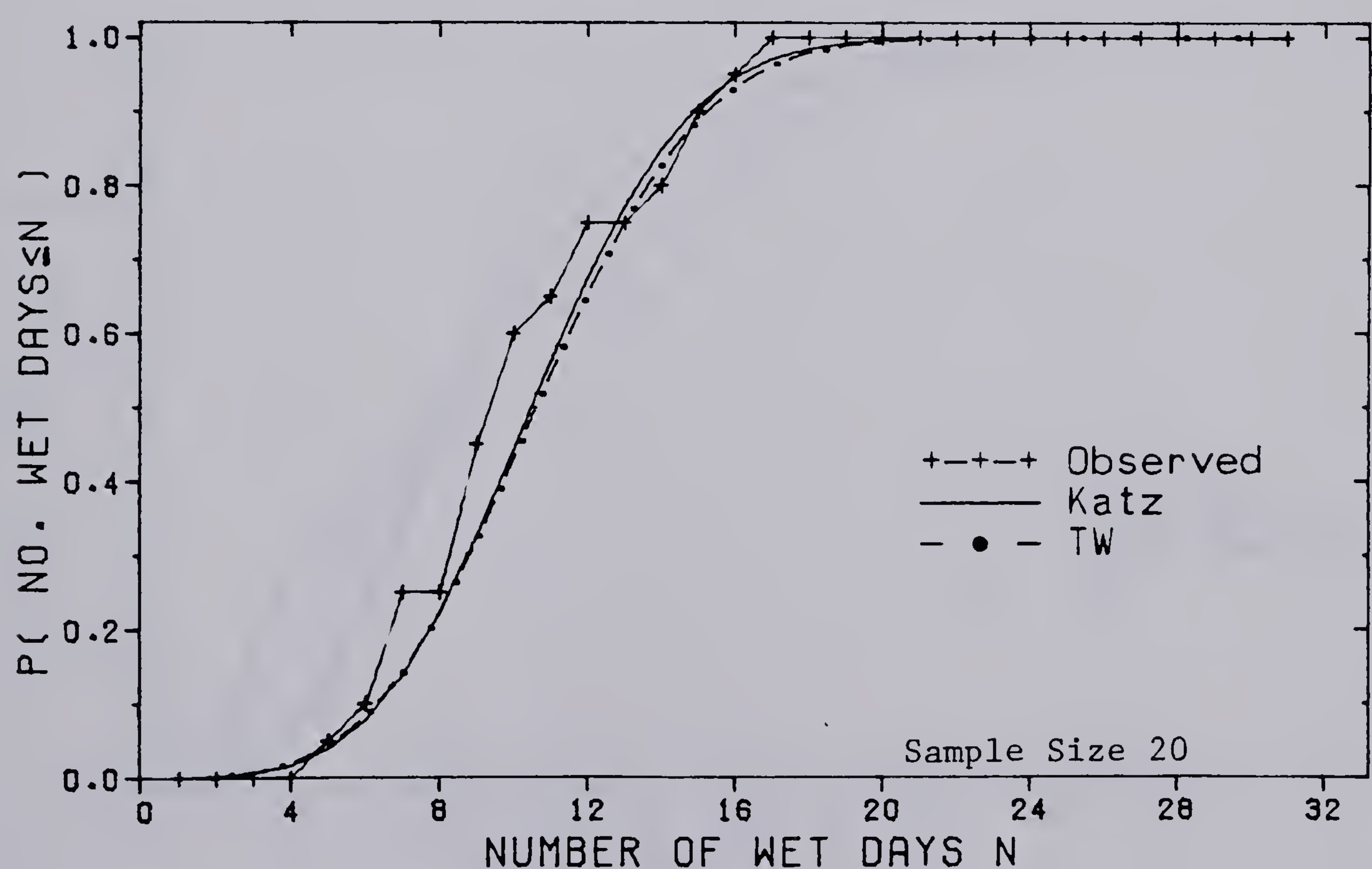
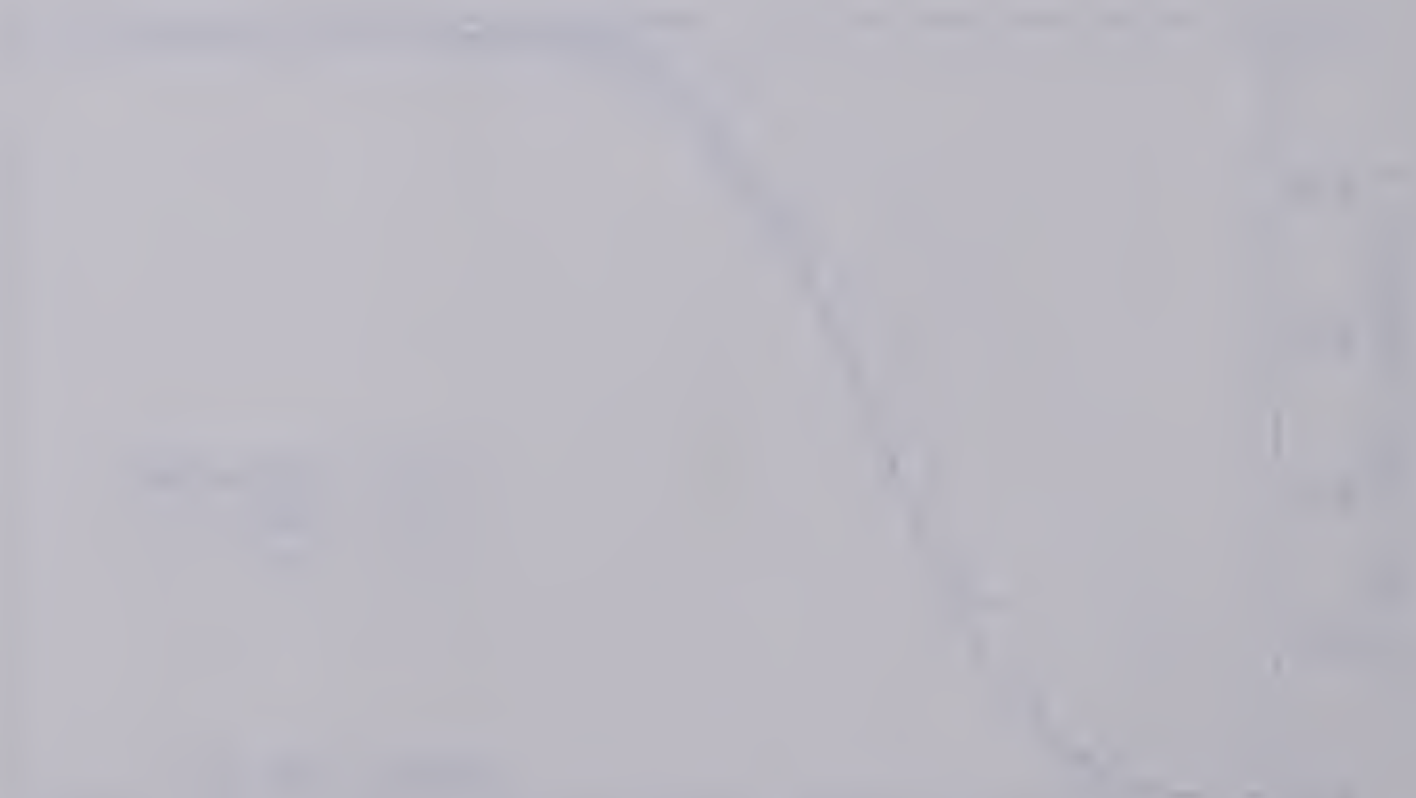


Figure 40. The theoretical distributions and the observed independent distribution for the number of wet days in July at Beaverlodge.



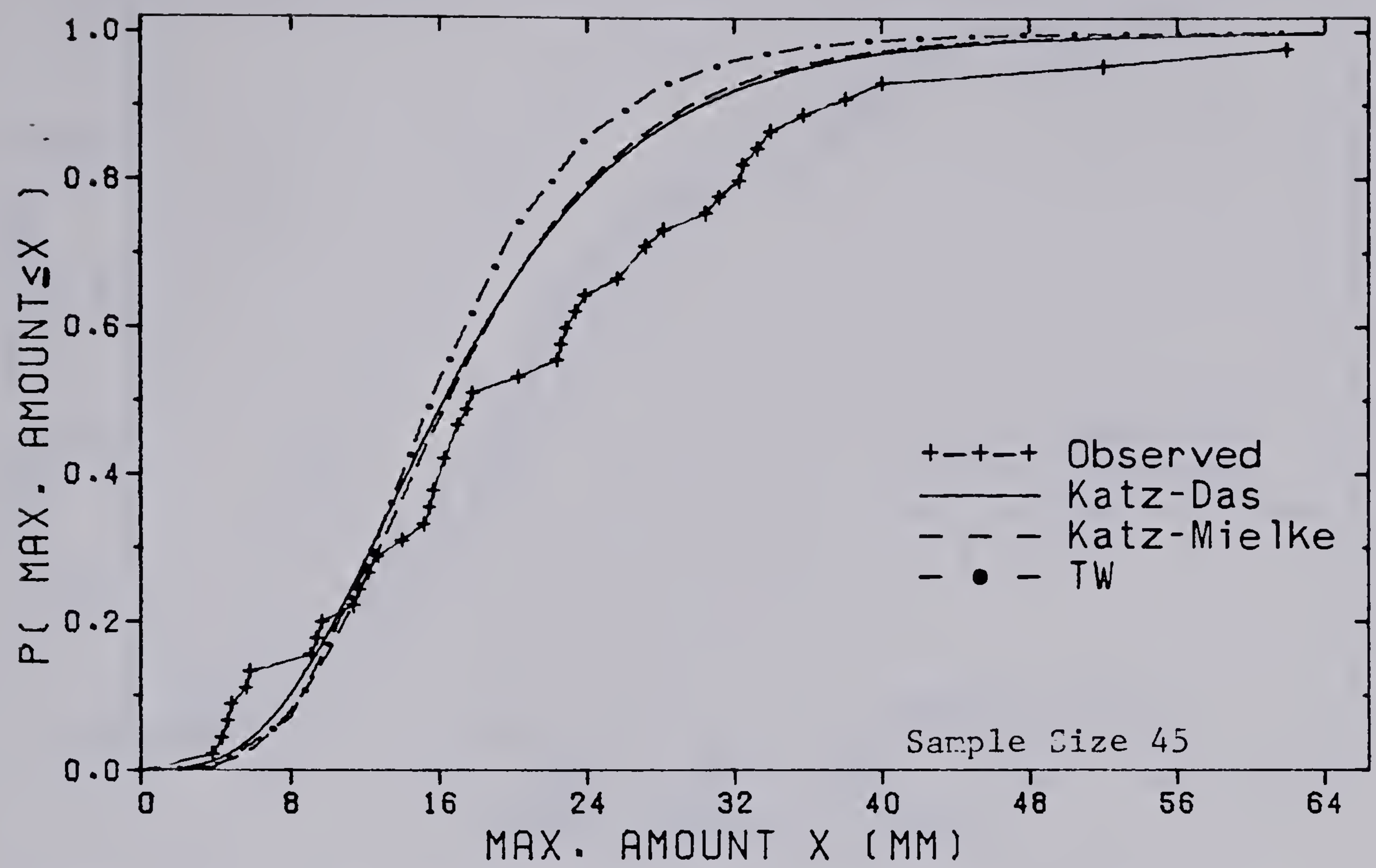


Figure 41. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in July at Beaverlodge.

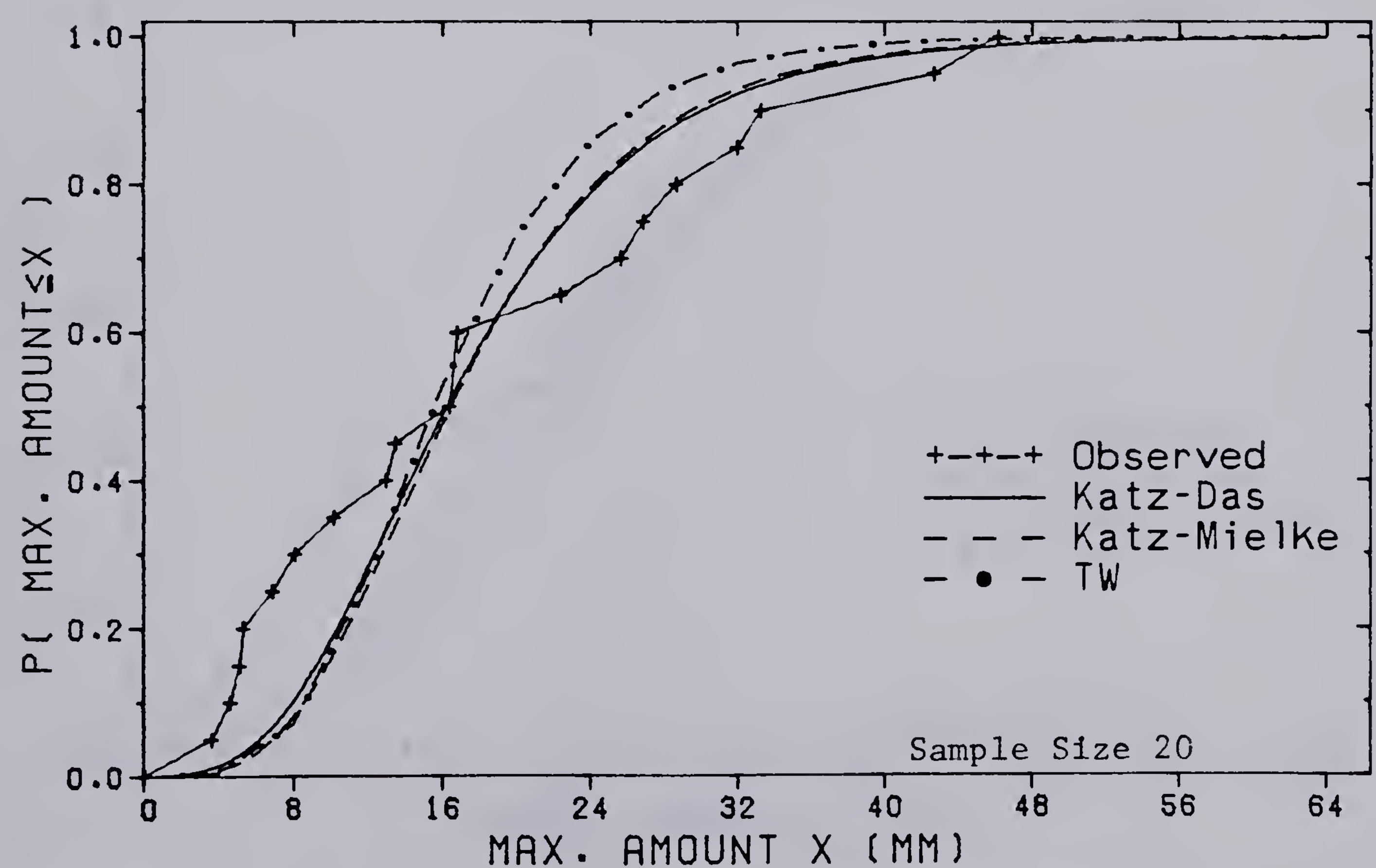


Figure 42. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in July at Beaverlodge.

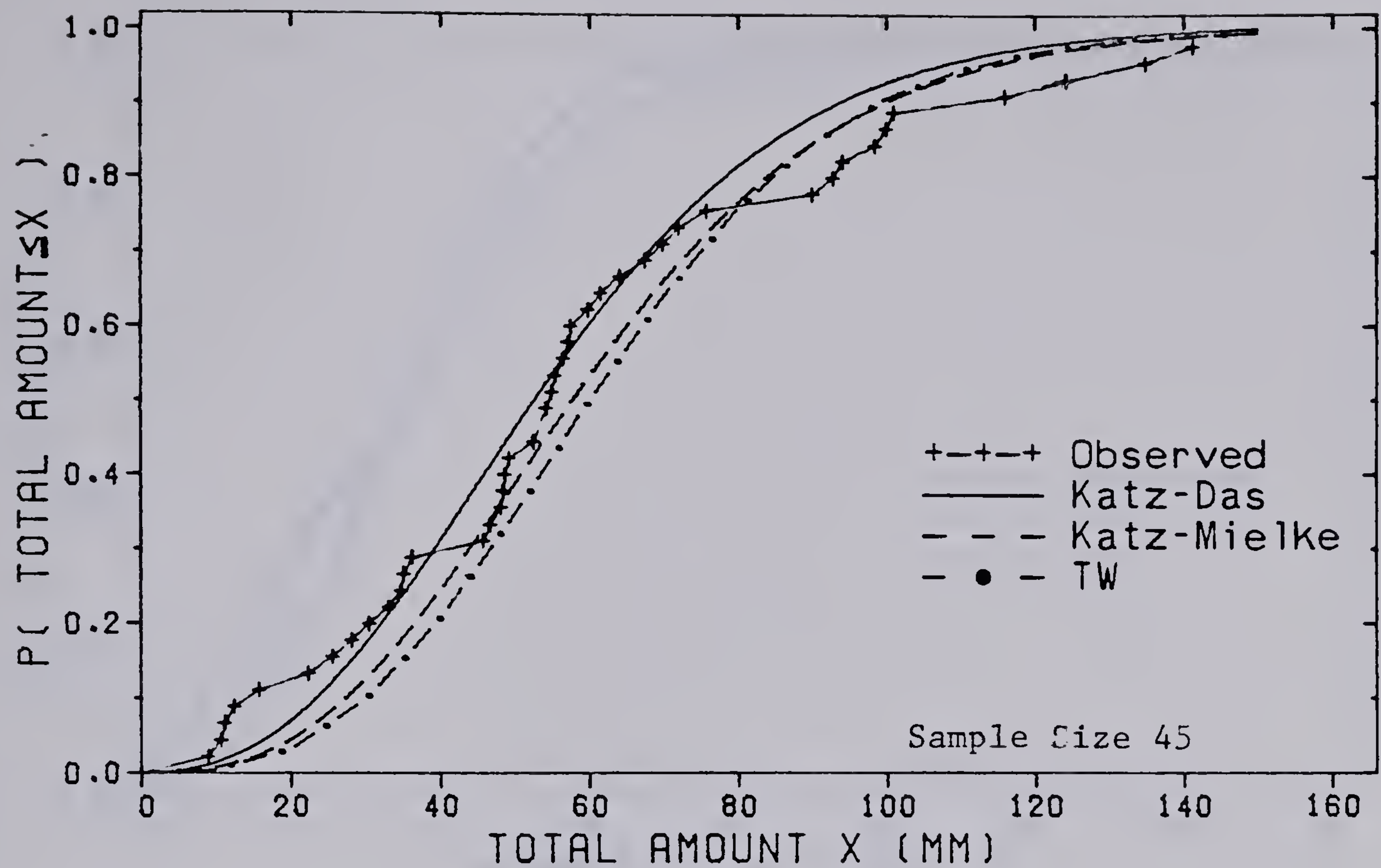


Figure 43. The theoretical distributions and the observed development distribution for the total amount of precipitation in July at Beaverlodge.

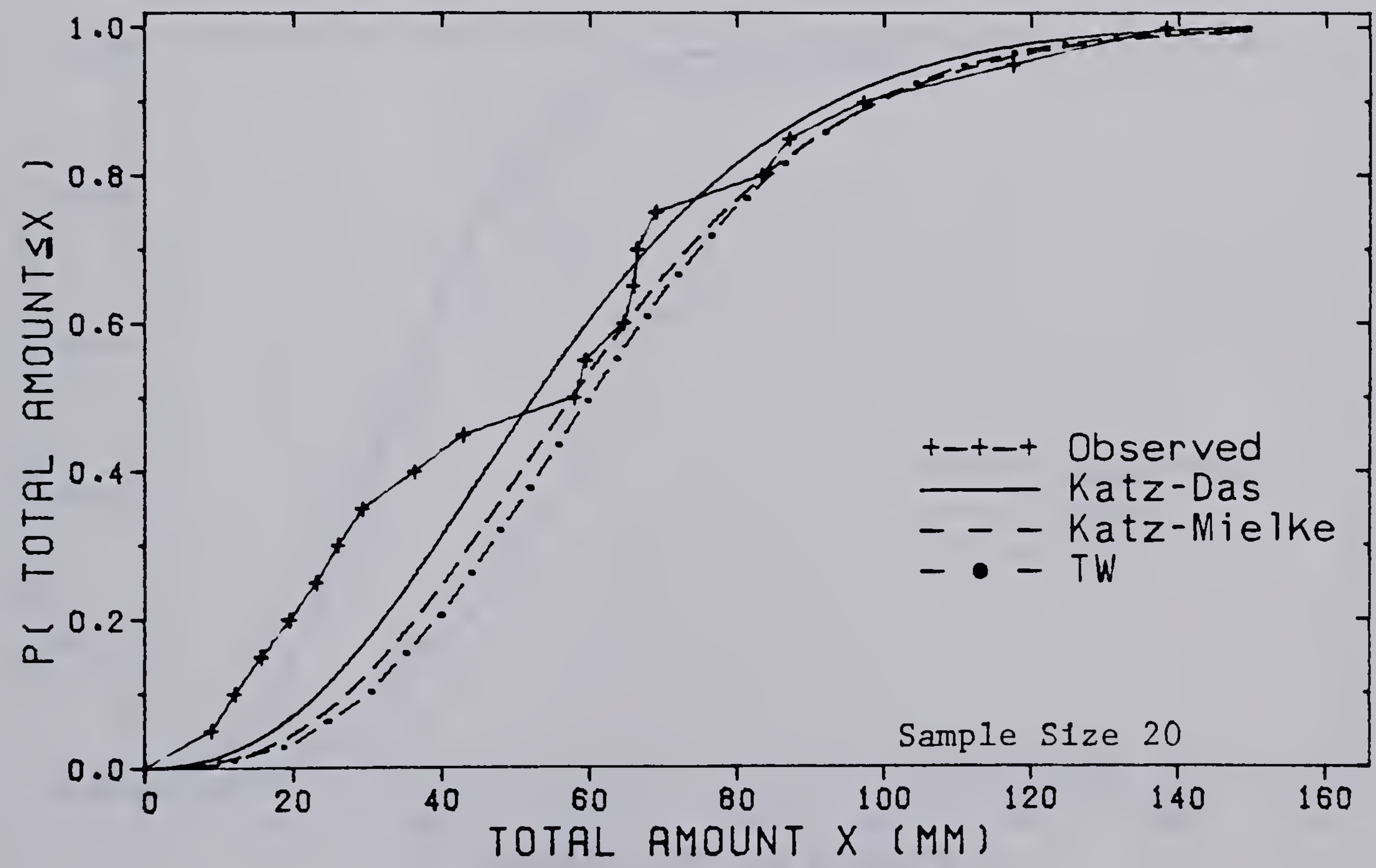


Figure 44. The theoretical distributions and the observed independent distribution for the total amount of precipitation in July at Beaverlodge.

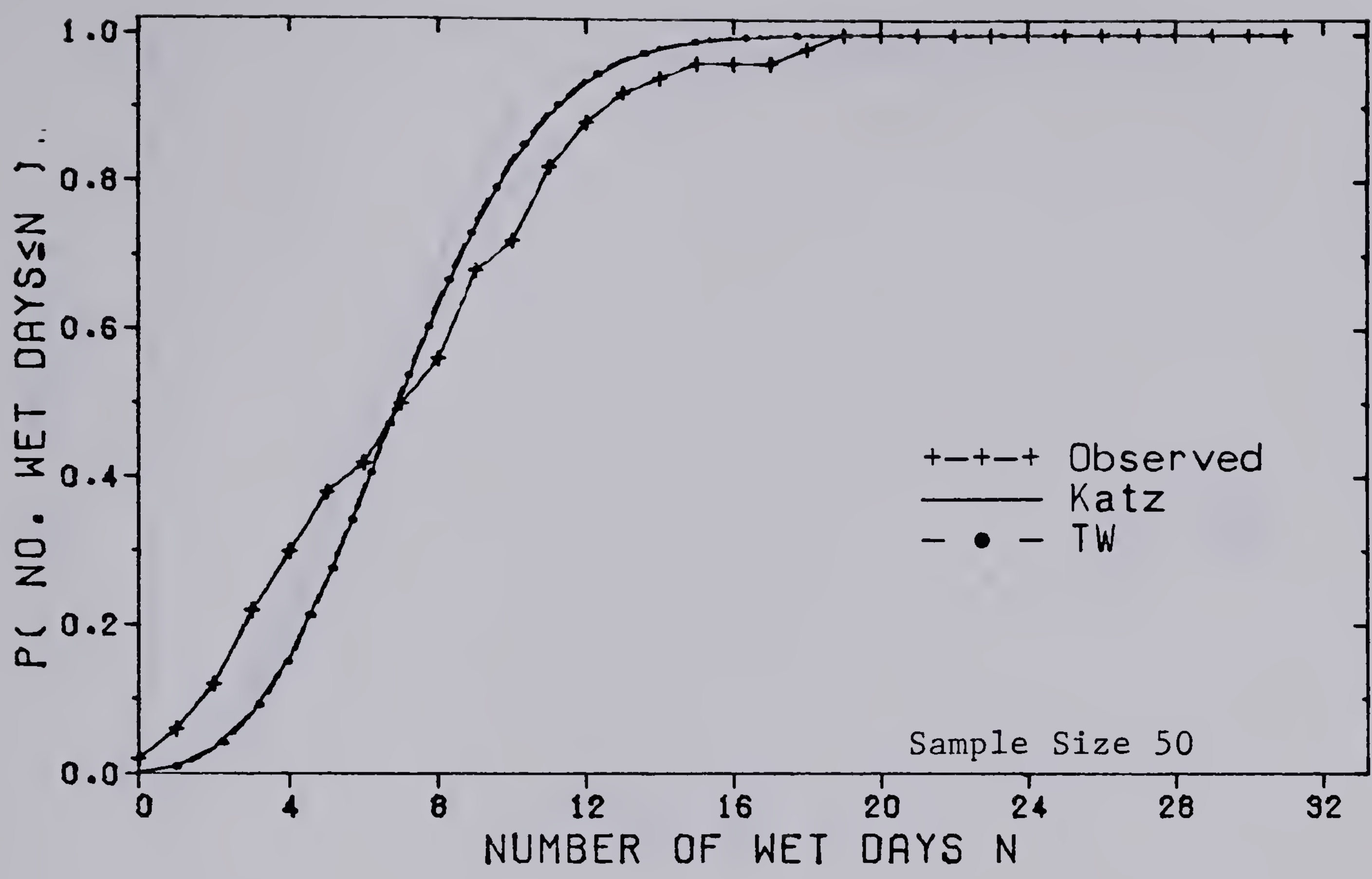


Figure 45. The theoretical distributions and the observed development distribution for the number of wet days in January at Edmonton.

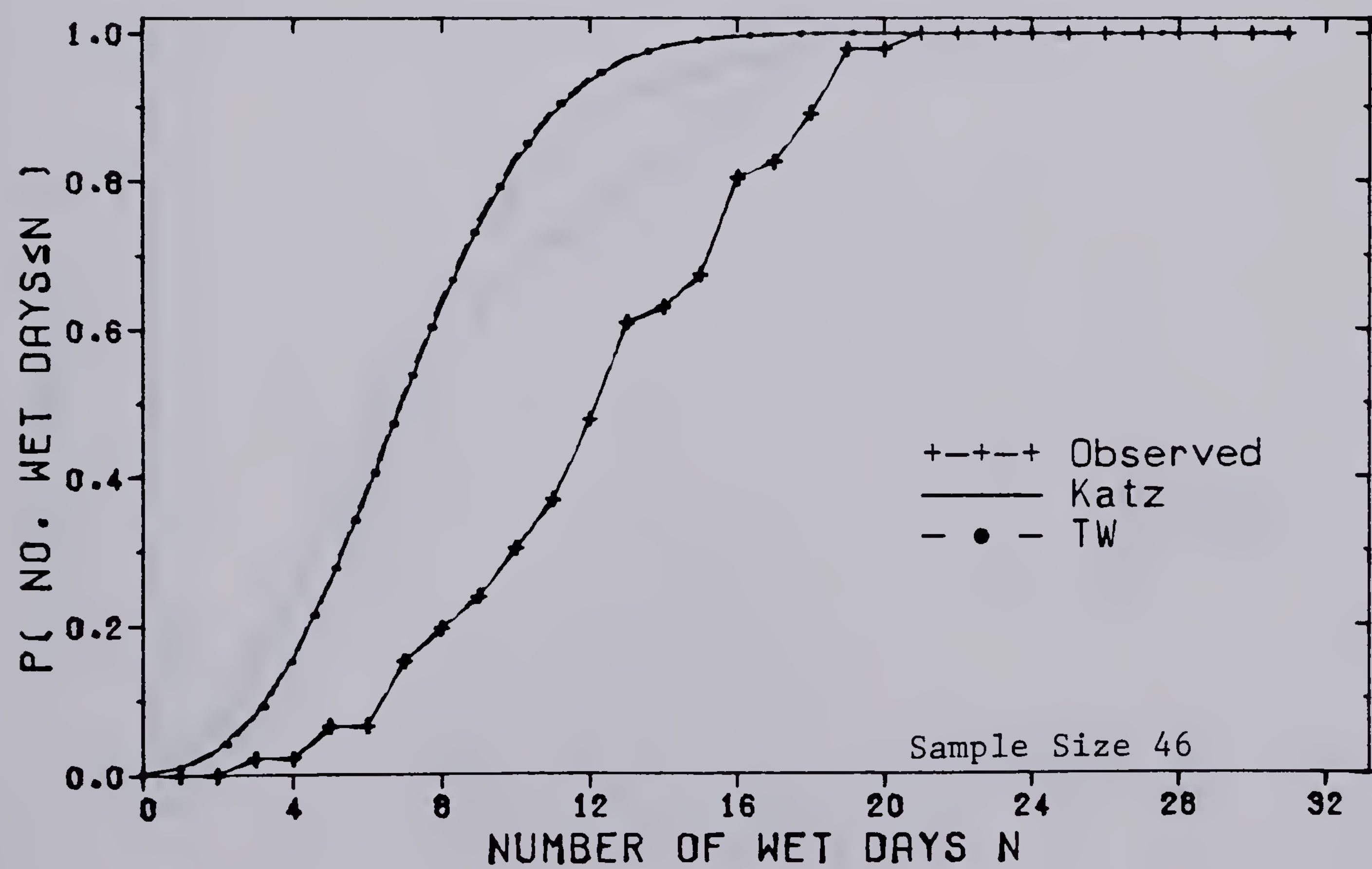


Figure 46. The theoretical distributions and the observed independent distribution for the number of wet days in January at Edmonton.

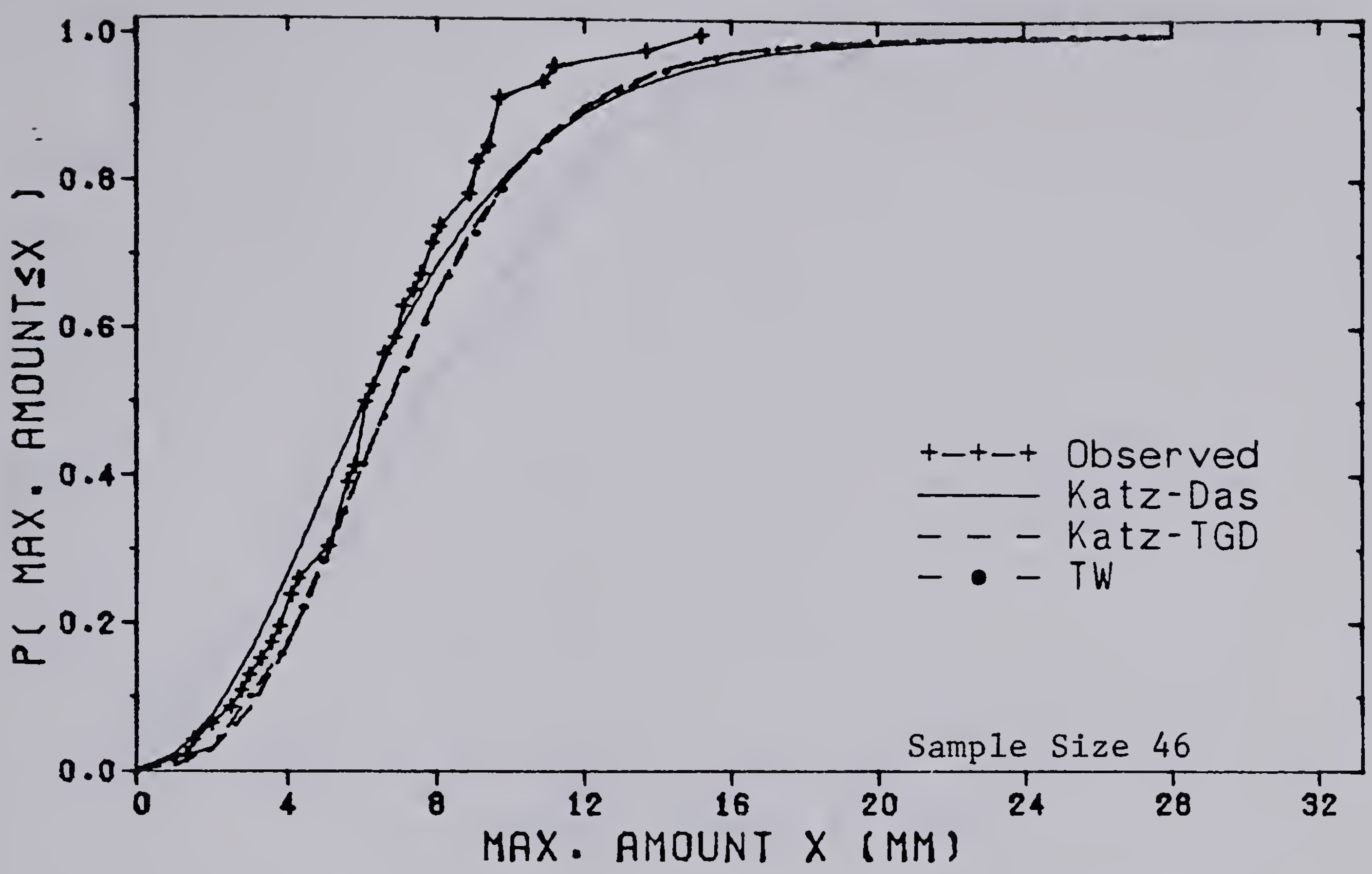


Figure 47. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in January at Edmonton.

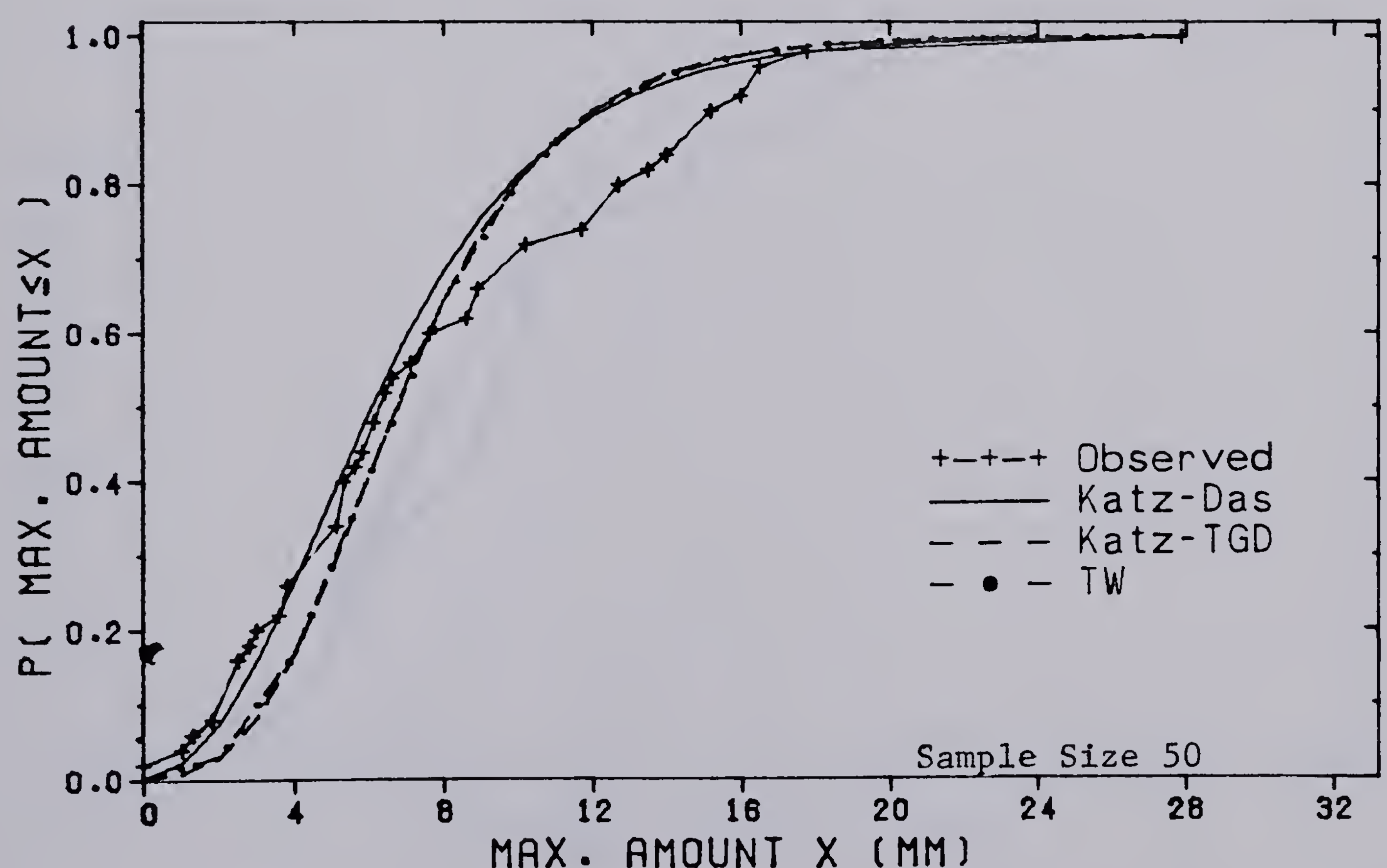


Figure 48. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in January at Edmonton.

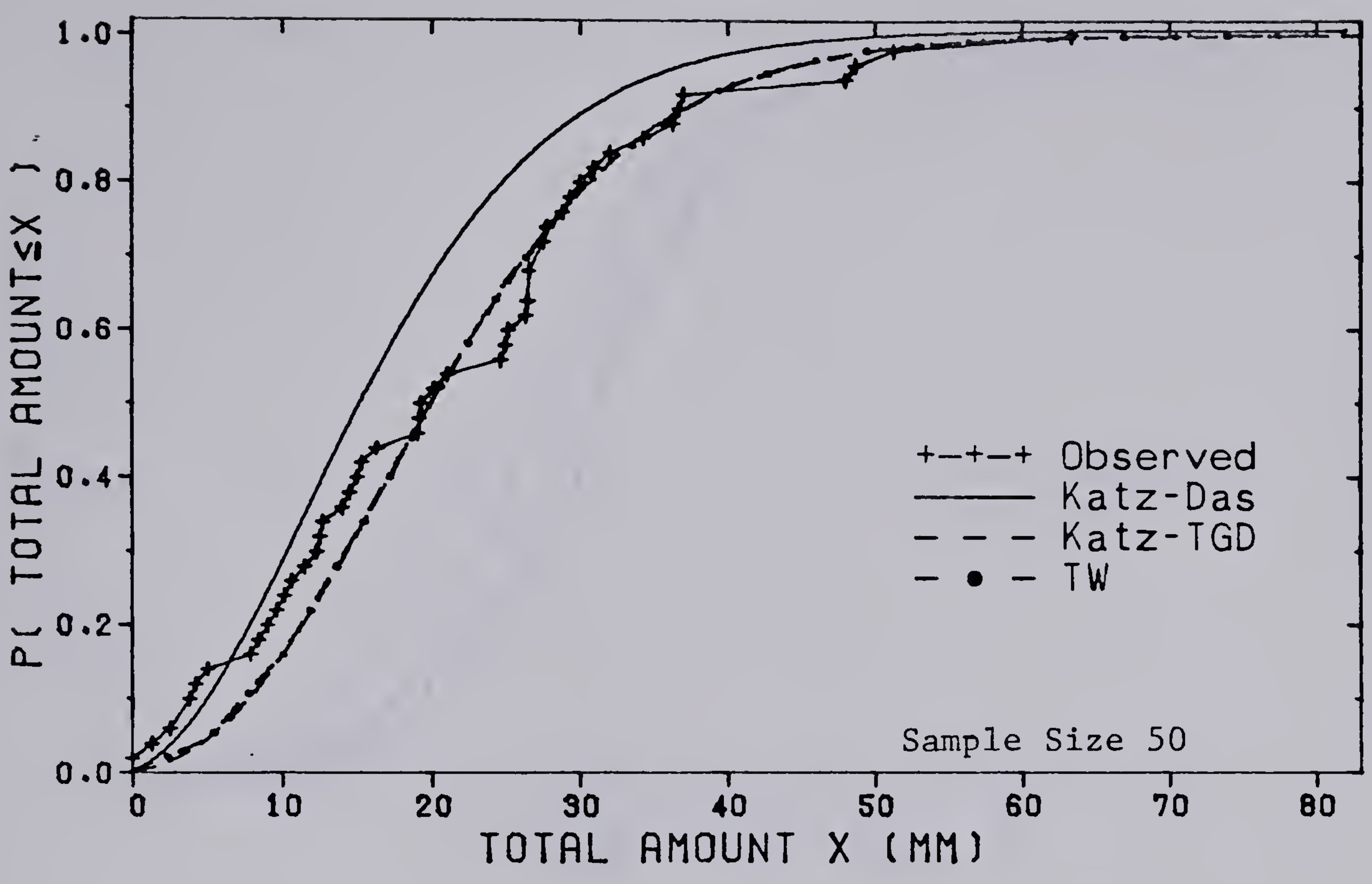


Figure 49. The theoretical distributions and the observed development distribution for the total amount of precipitation in January at Edmonton.

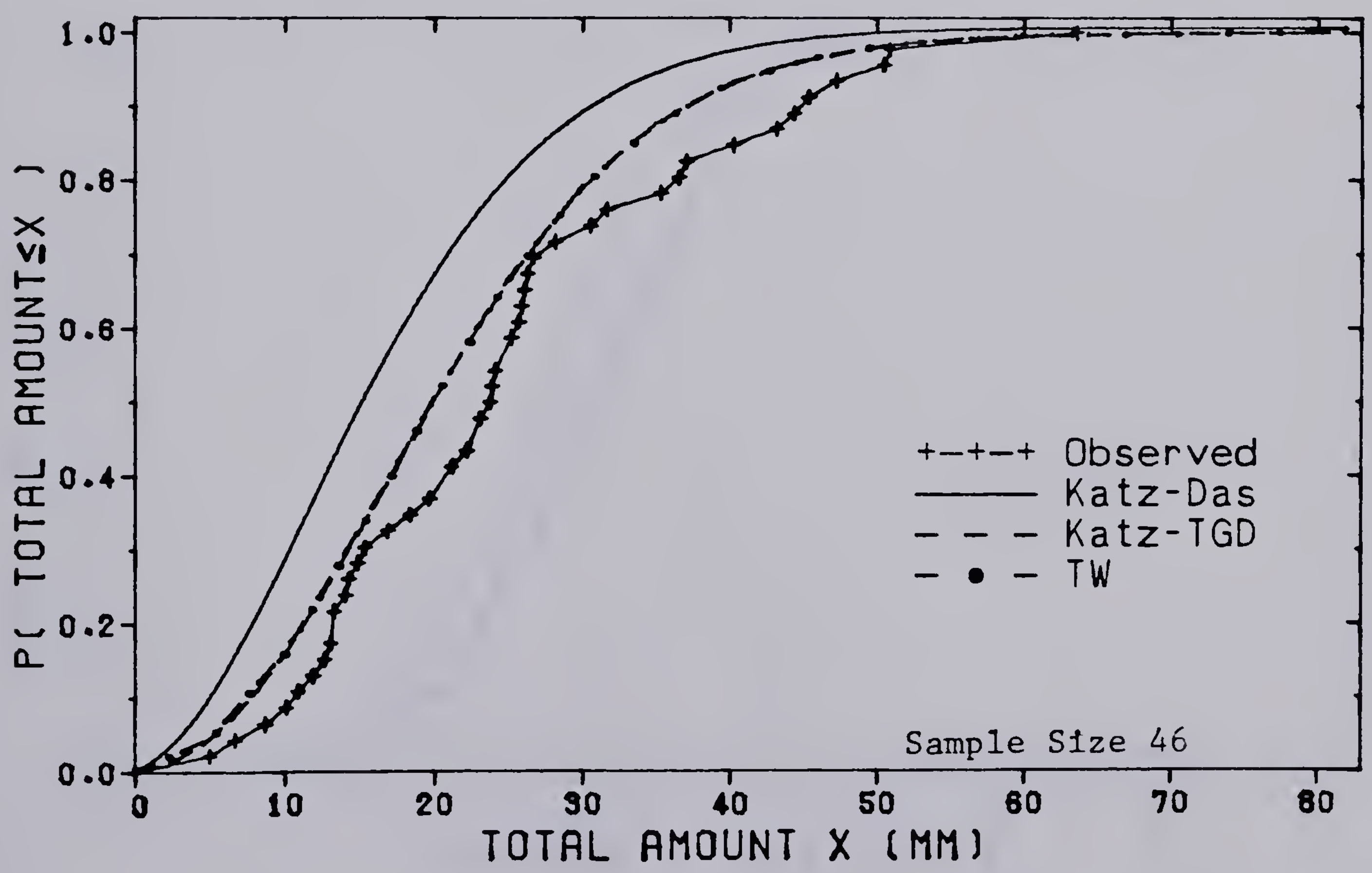


Figure 50. The theoretical distributions and the observed independent distribution for the total amount of precipitation in January at Edmonton.

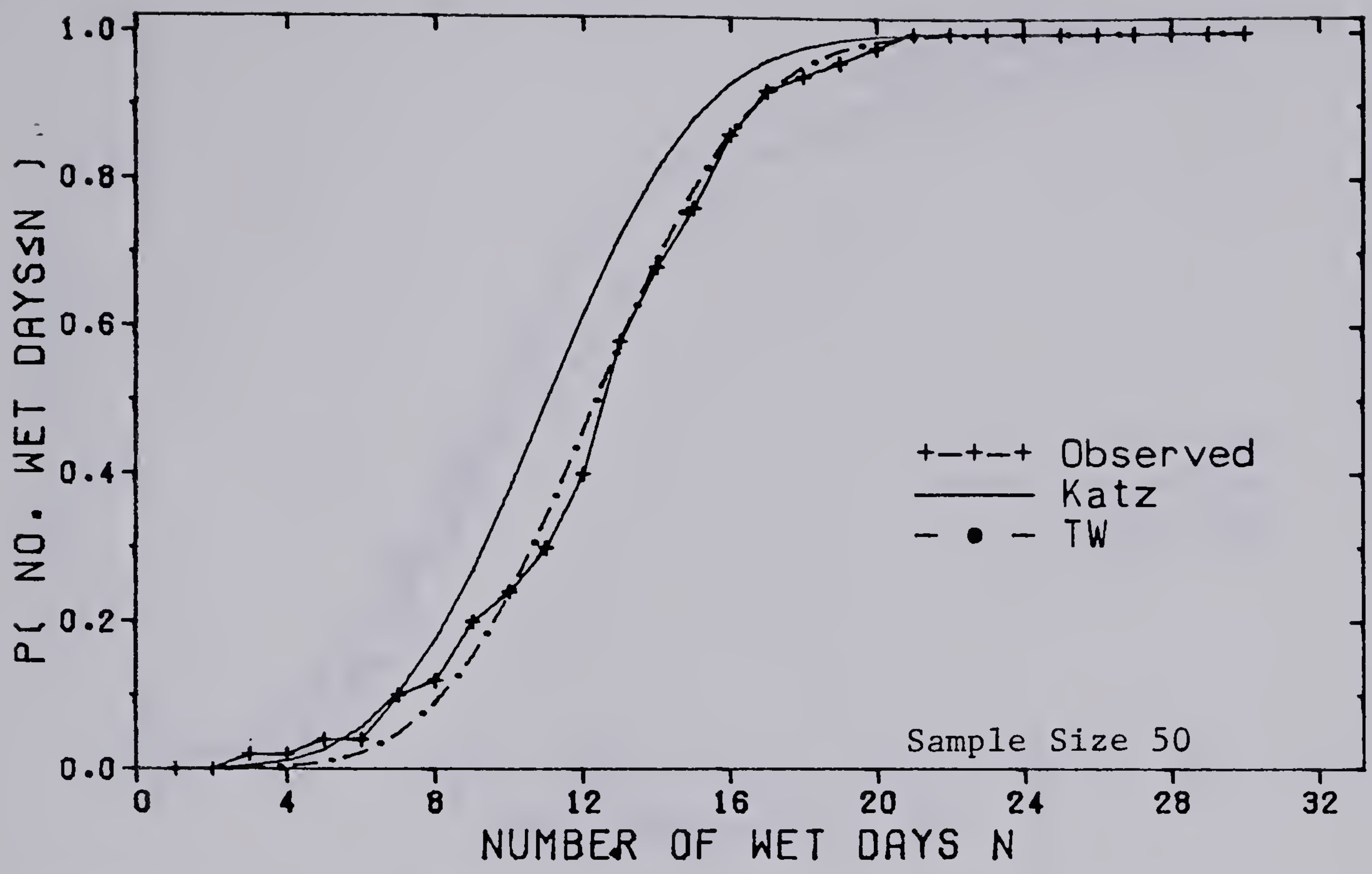


Figure 51. The theoretical distributions and the observed development distribution for the number of wet days in June at Edmonton.

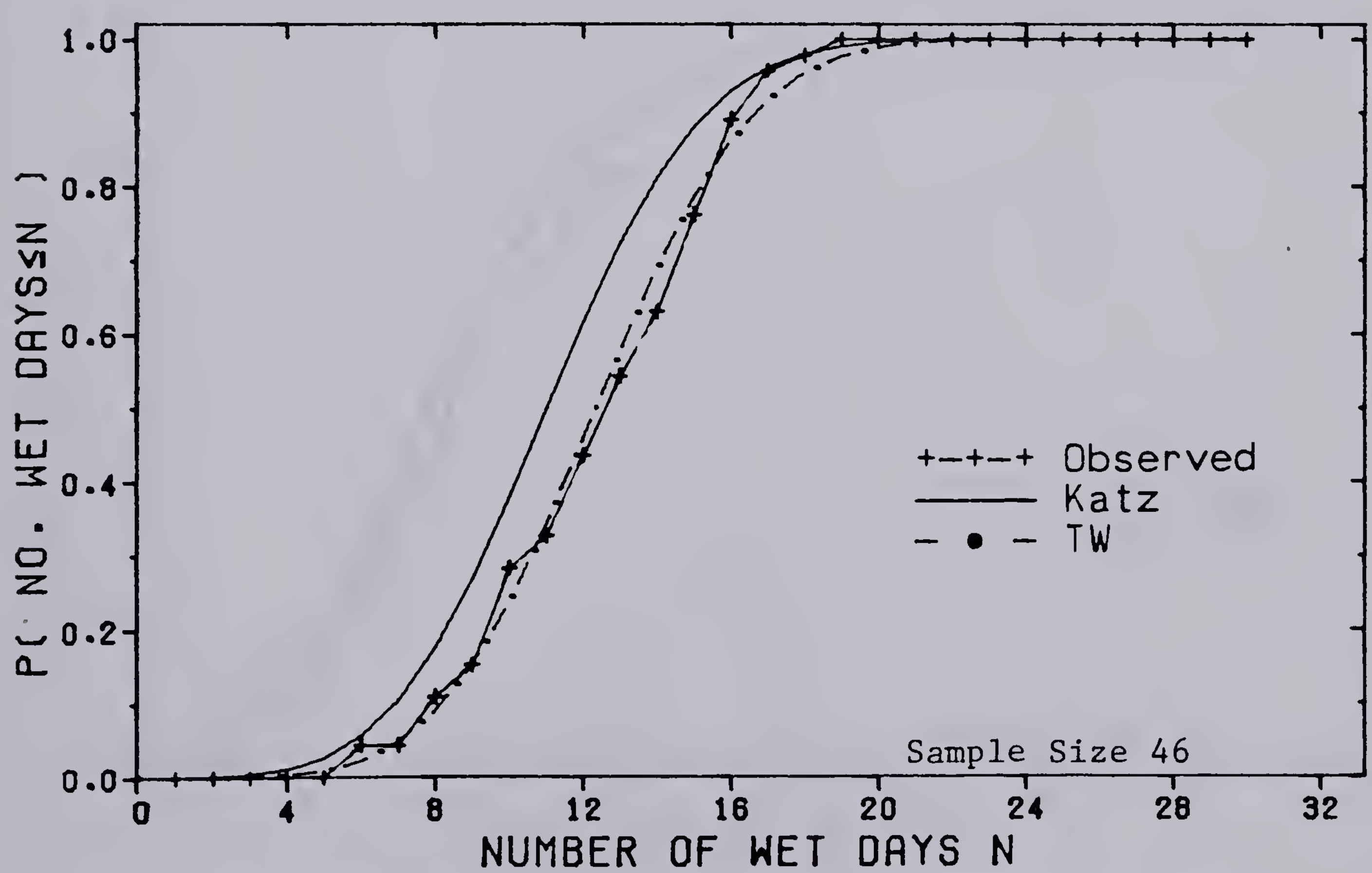


Figure 52. The theoretical distributions and the observed independent distribution for the number of wet days in June at Edmonton.

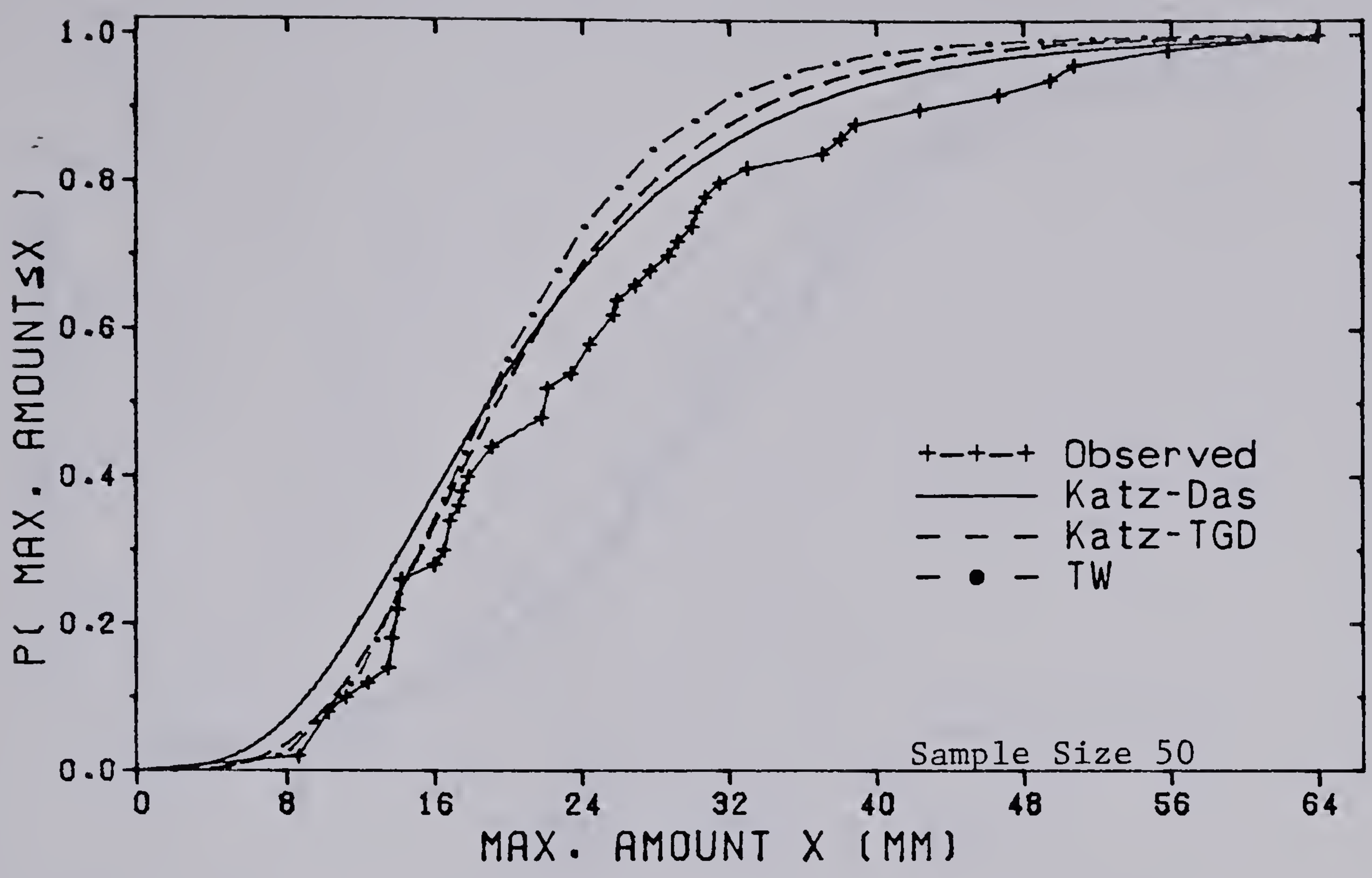


Figure 53. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in June at Edmonton.

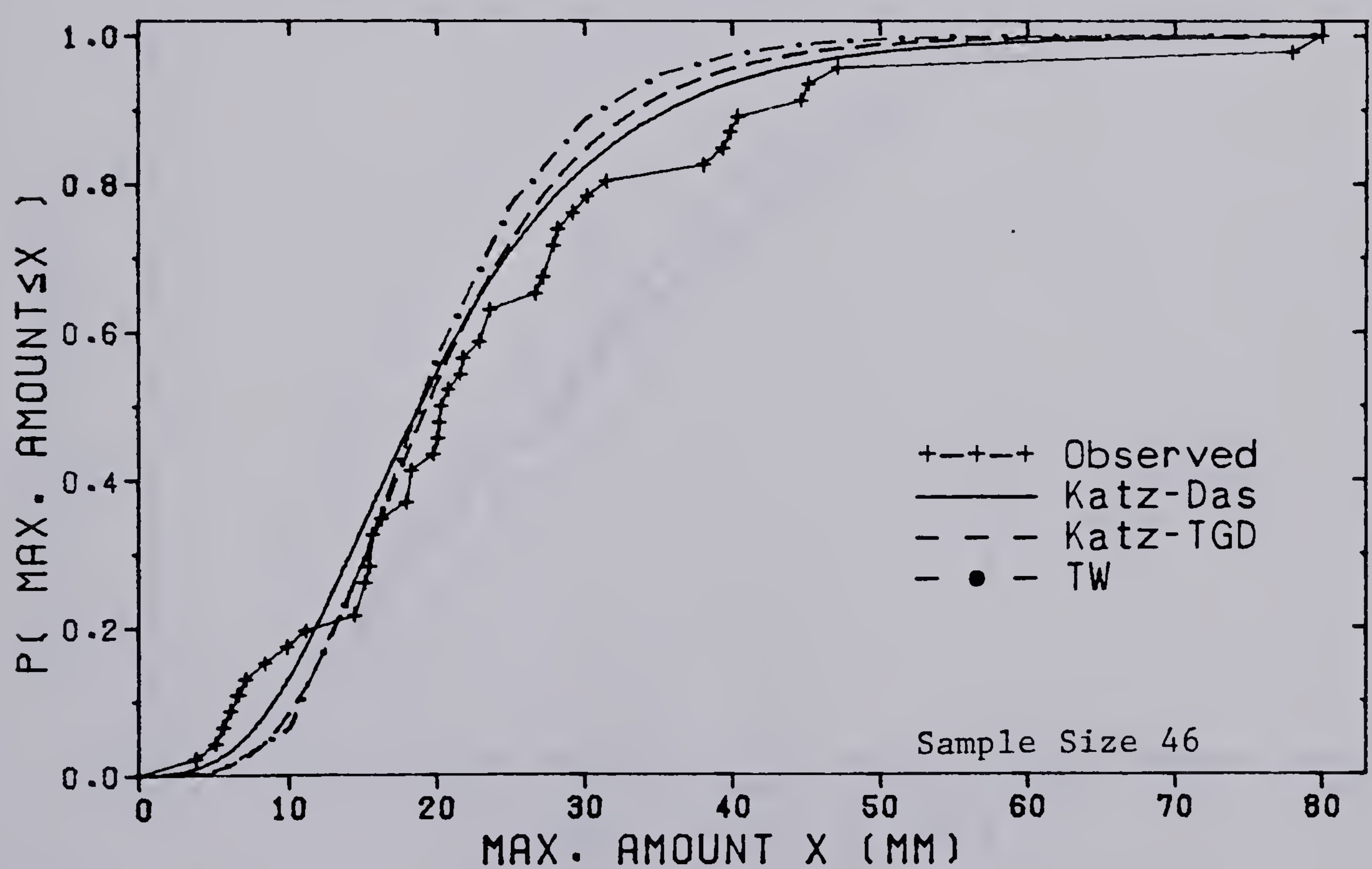


Figure 54. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in June at Edmonton.

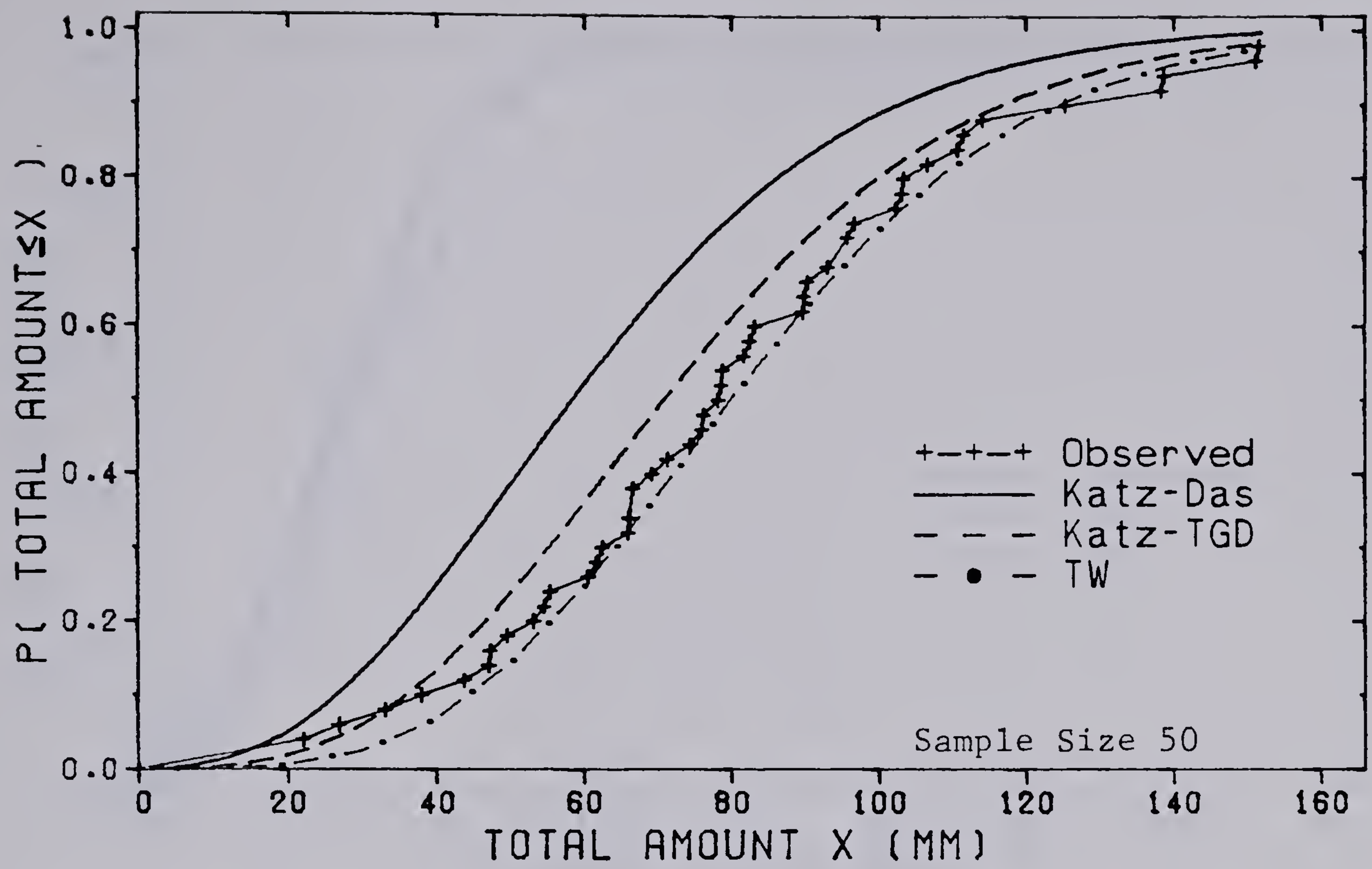


Figure 55. The theoretical distributions and the observed development distribution for the total amount of precipitation in June at Edmonton.

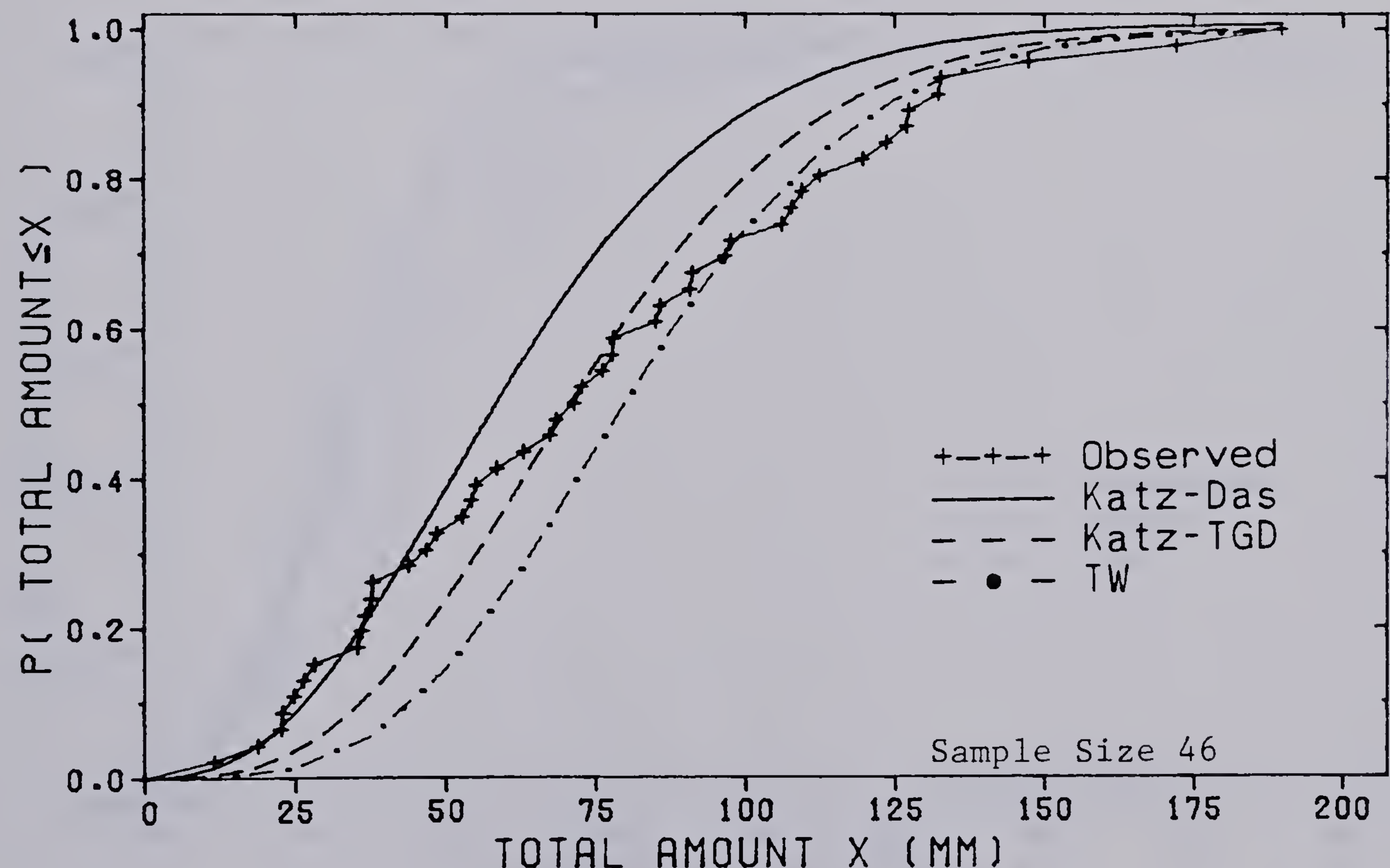


Figure 56. The theoretical distributions and the observed independent distribution for the total amount of precipitation in June at Edmonton.

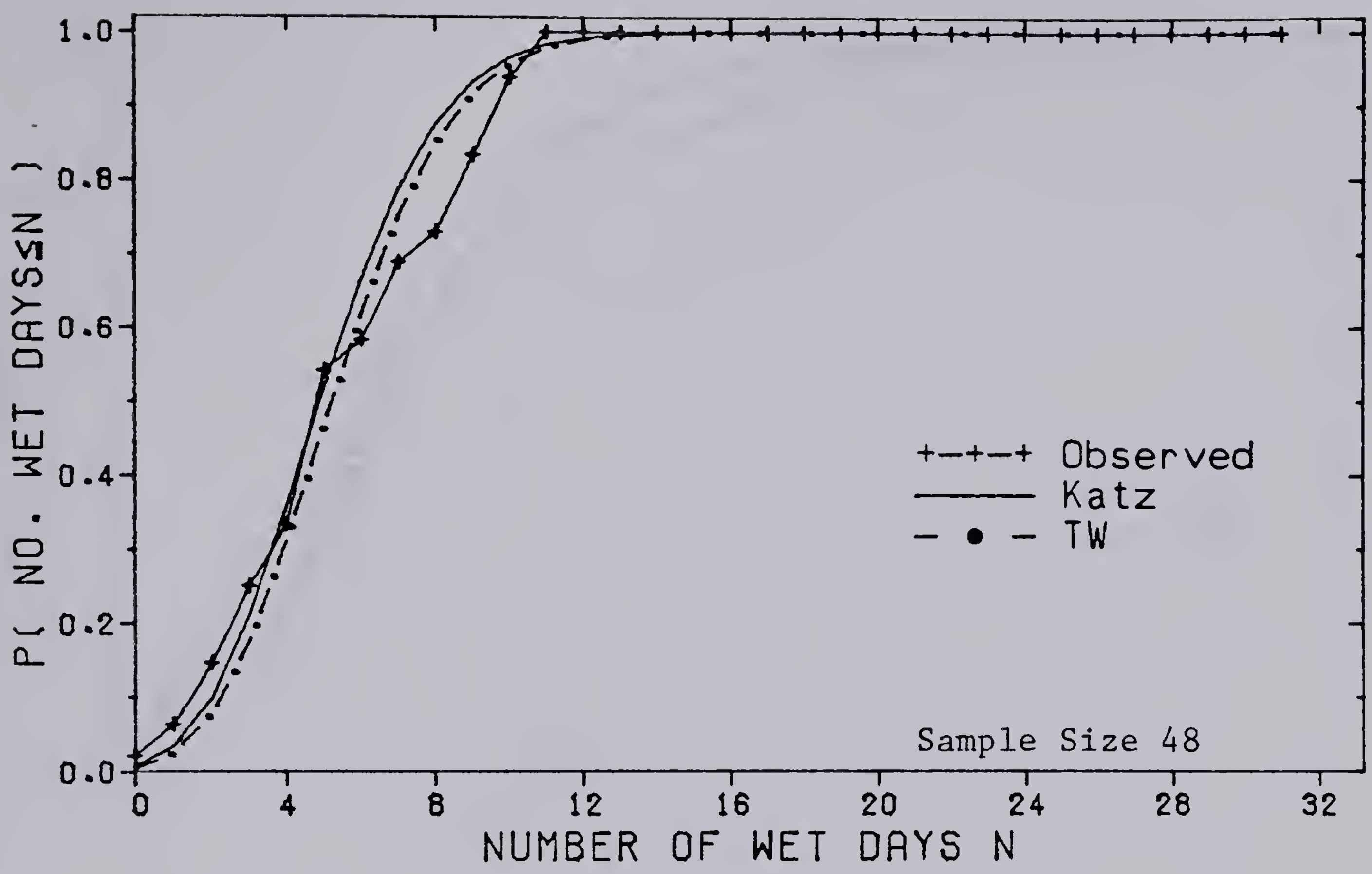


Figure 57. The theoretical distributions and the observed development distribution for the number of wet days in March at Medicine Hat.

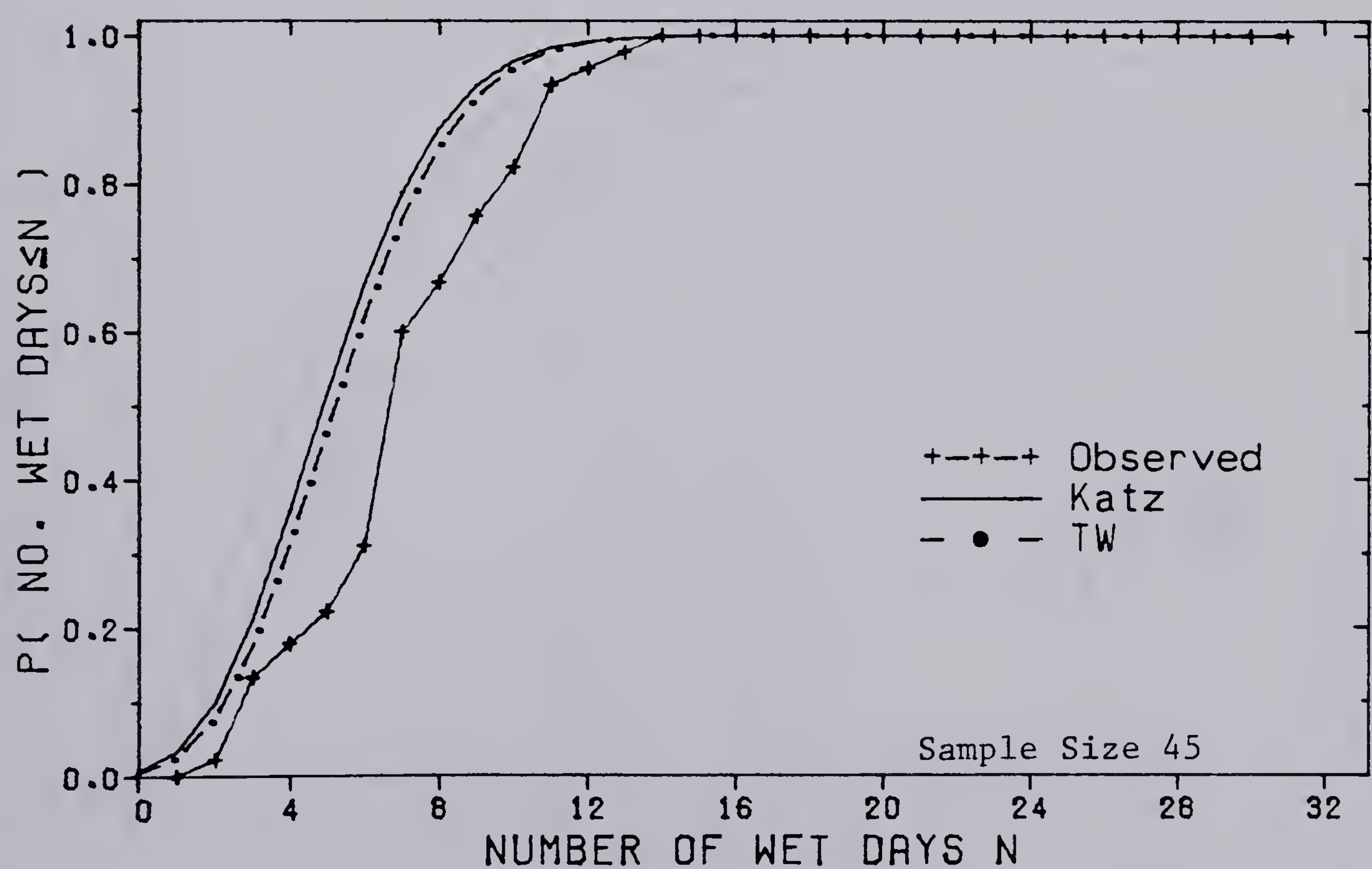


Figure 58. The theoretical distributions and the observed independent distribution for the number of wet days in March at Medicine Hat.

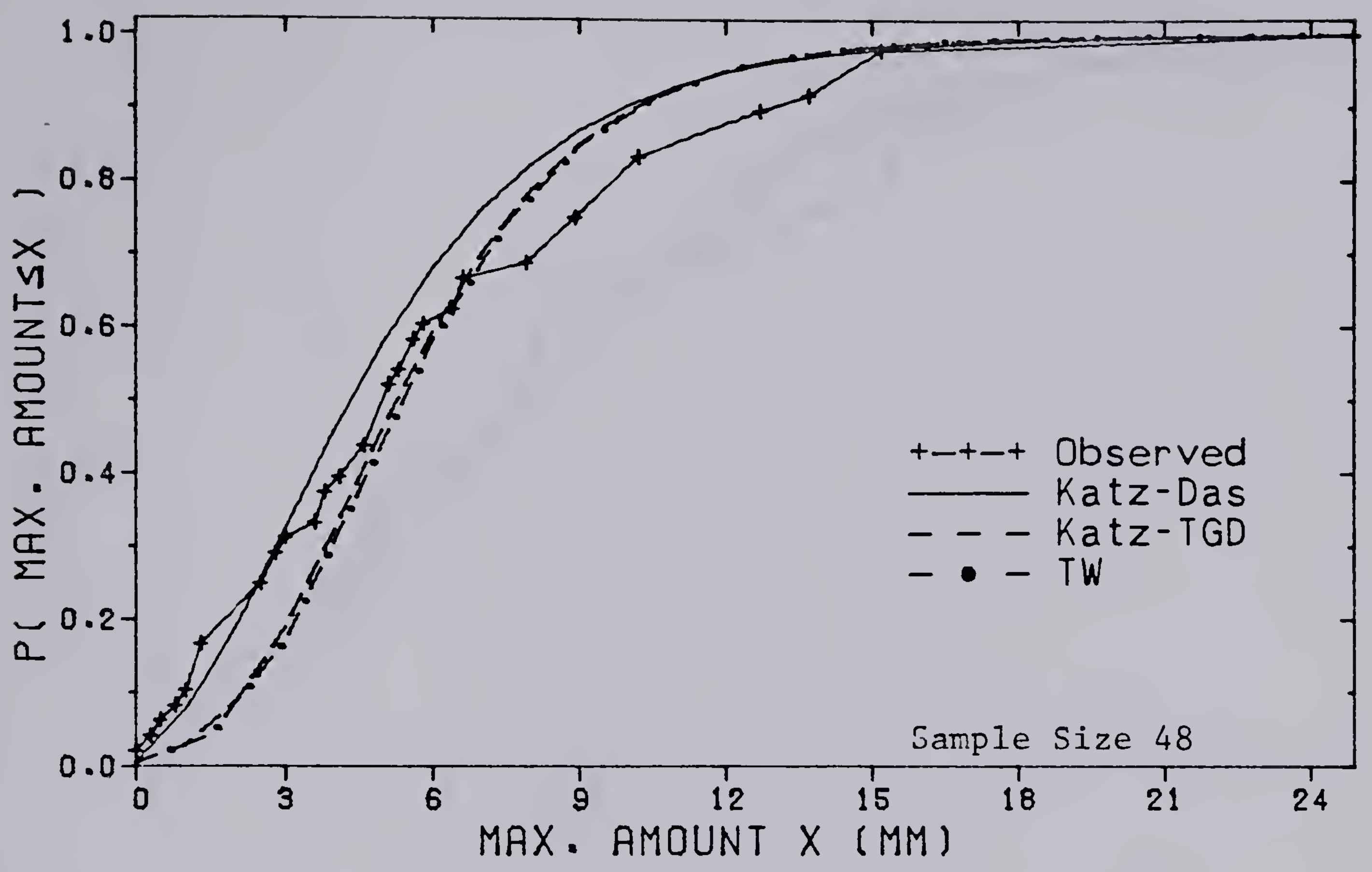


Figure 59. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in March at Medicine Hat.

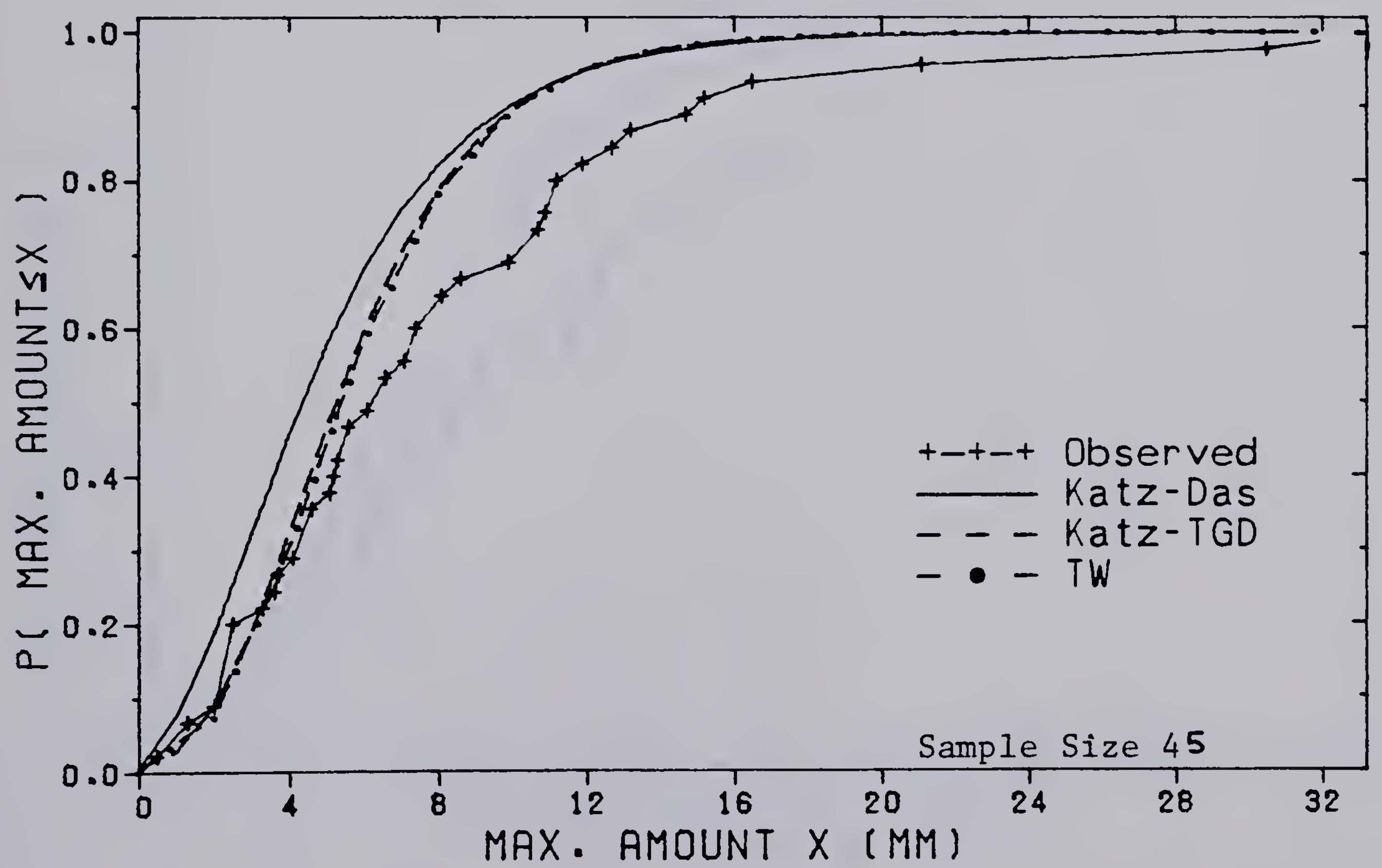


Figure 60. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in March at Medicine Hat.

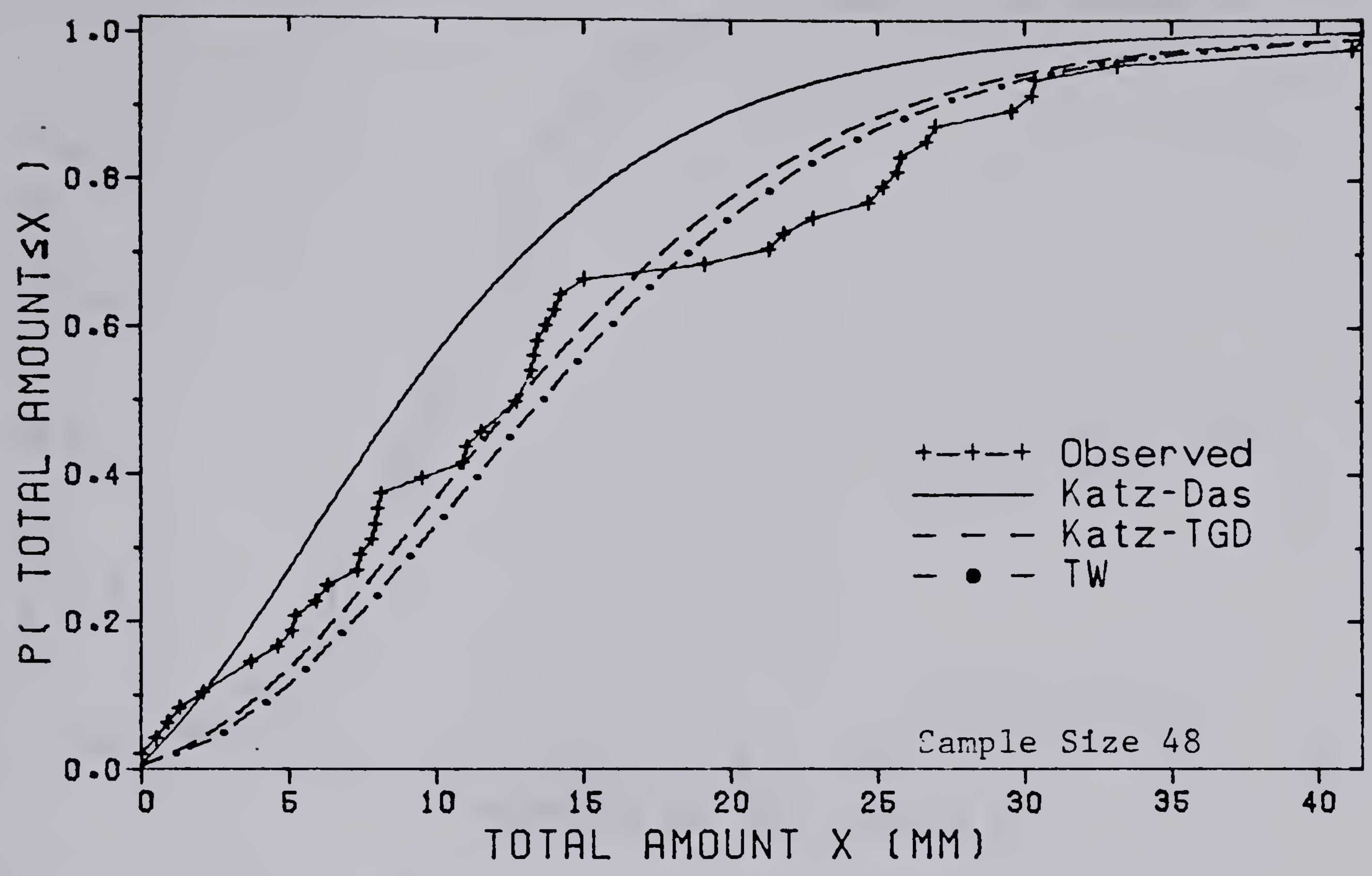


Figure 61. The theoretical distributions and the observed development distribution for the total amount of precipitation in March at Medicine Hat.

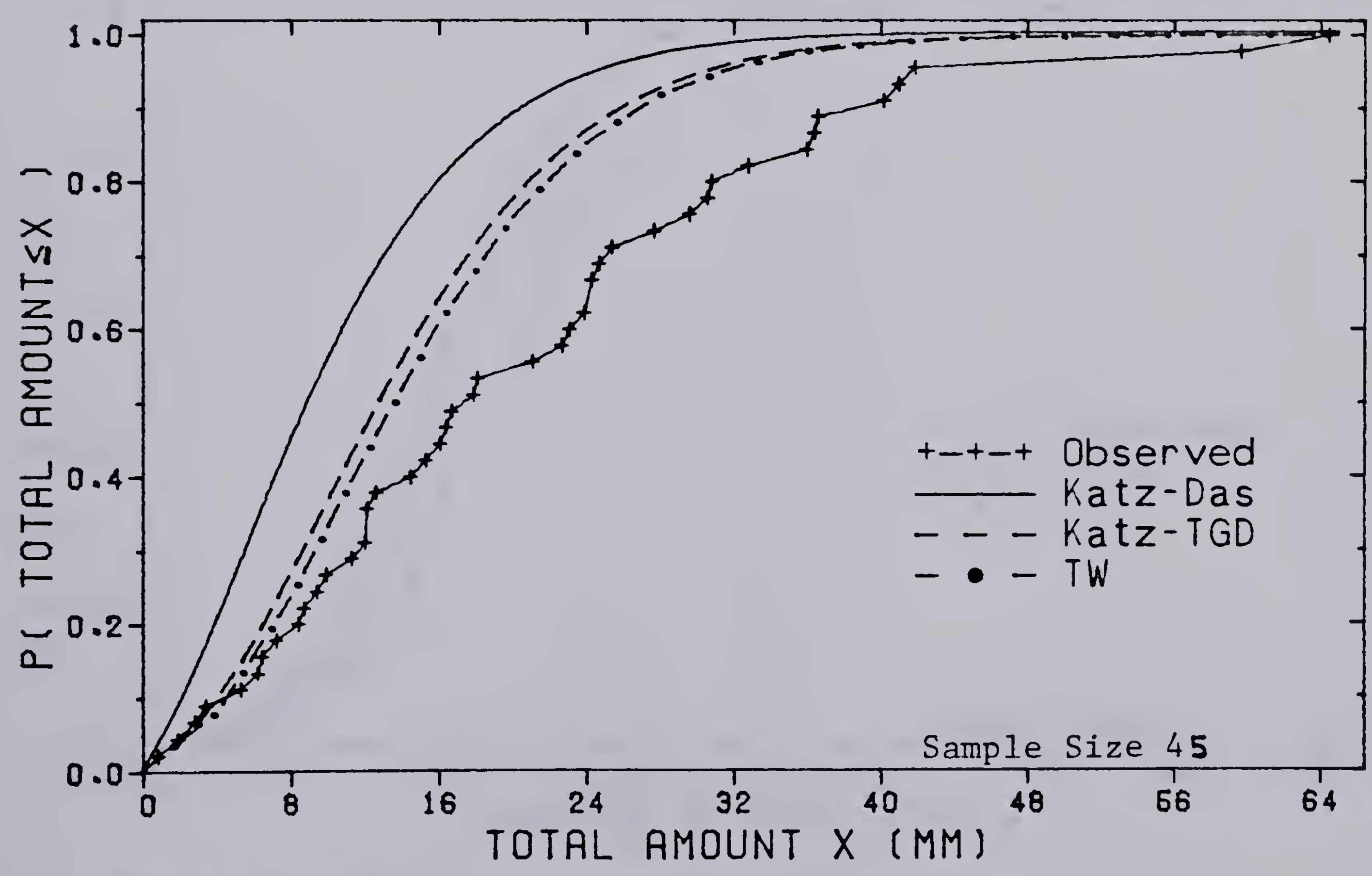


Figure 62. The theoretical distributions and the observed independent distribution for the total amount of precipitation in March at Medicine Hat.

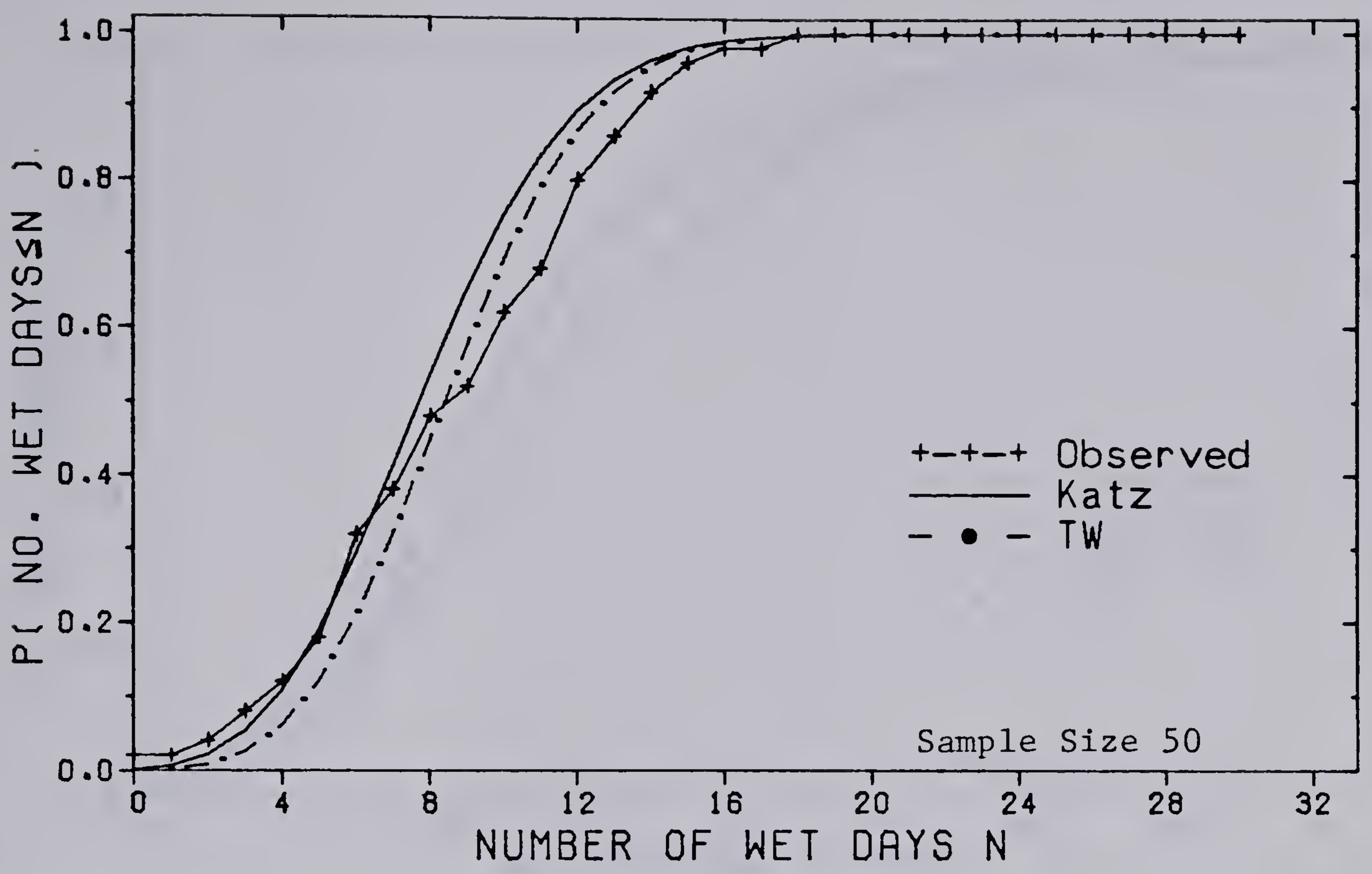


Figure 63. The theoretical distributions and the observed development distribution for the number of wet days in June at Medicine Hat.

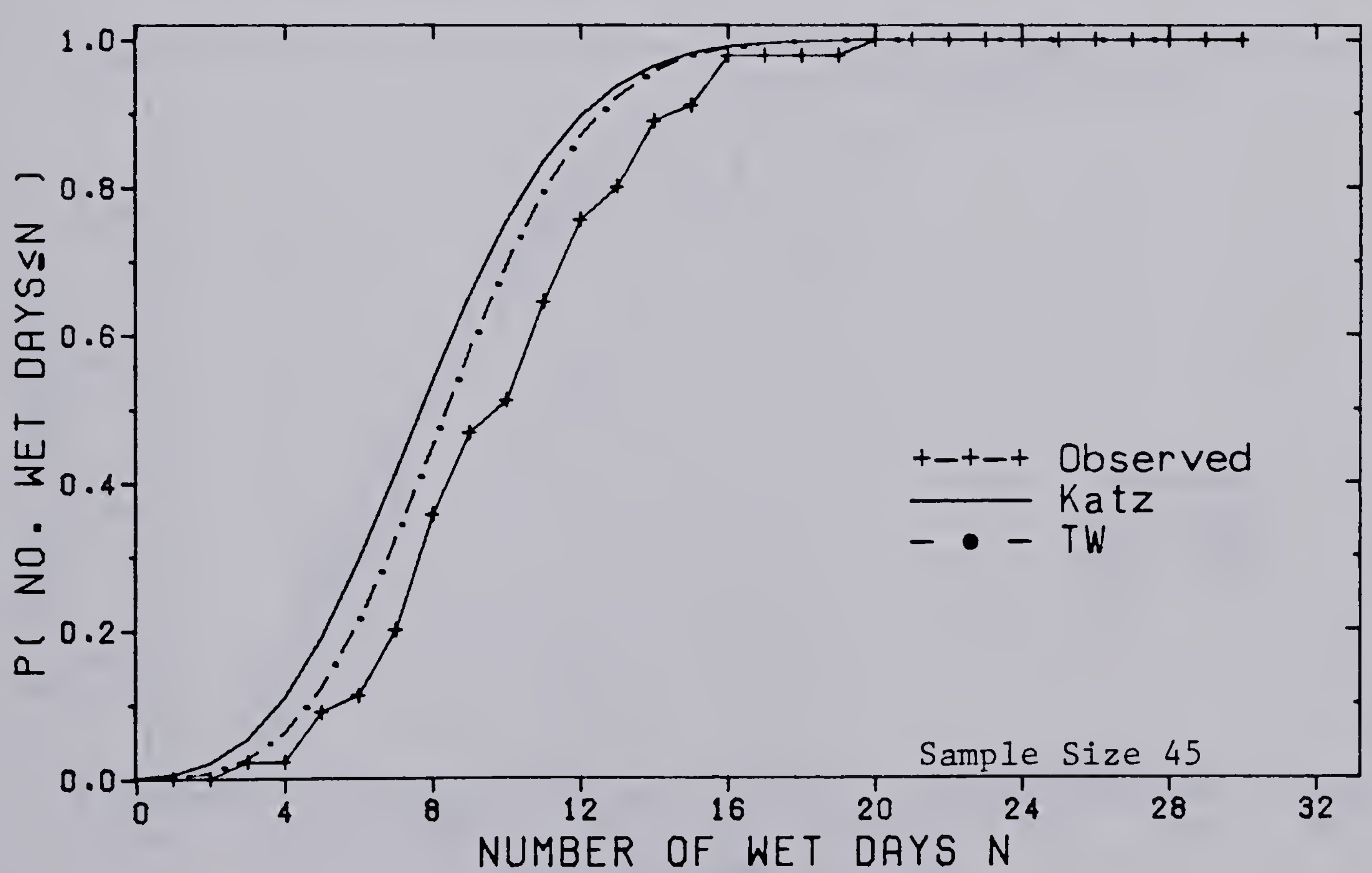
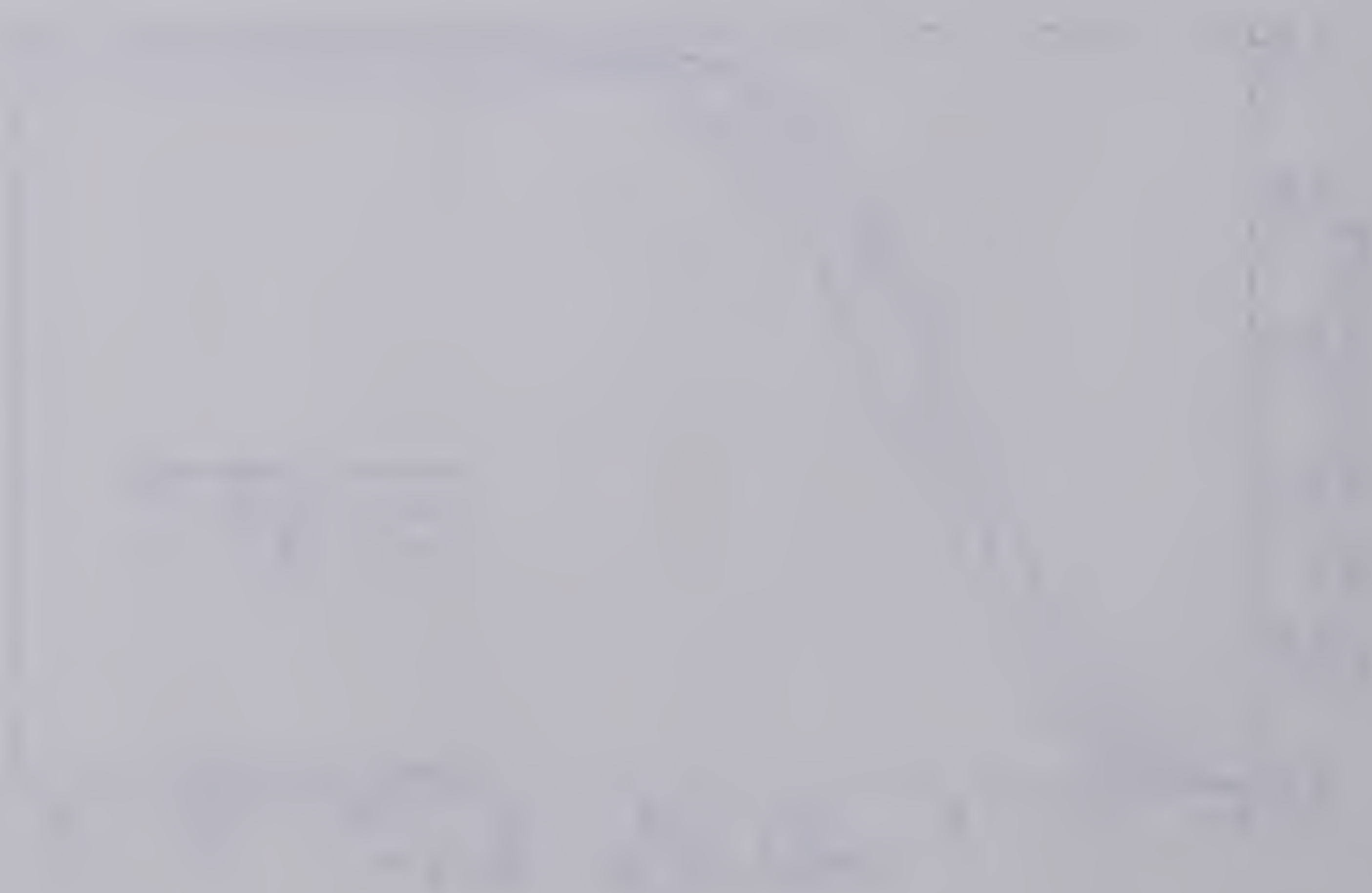


Figure 64. The theoretical distributions and the observed independent distribution for the number of wet days in June at Medicine Hat.



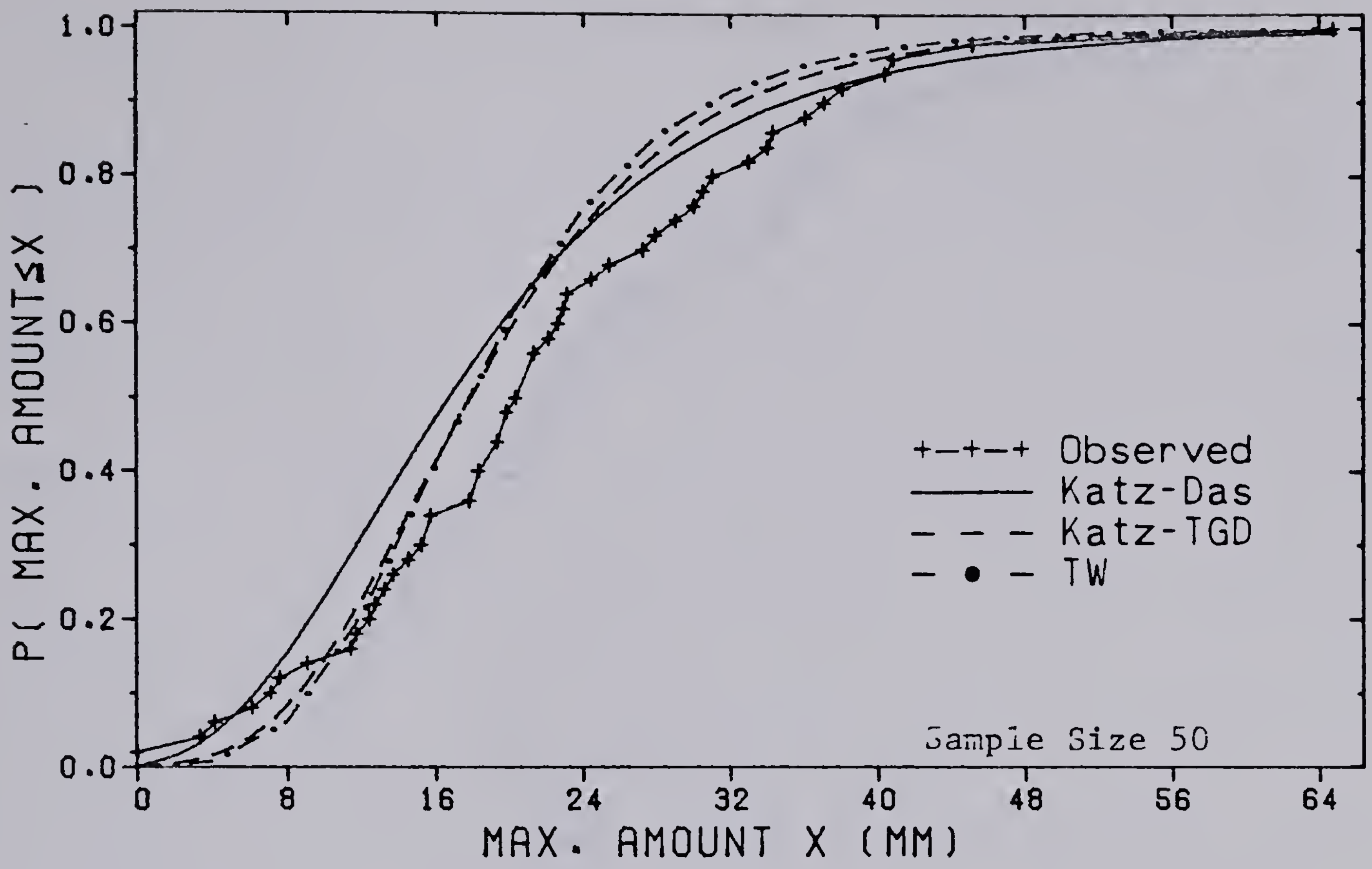


Figure 65. The theoretical distributions and the observed development distribution for the maximum daily amount of precipitation in June at Medicine Hat.

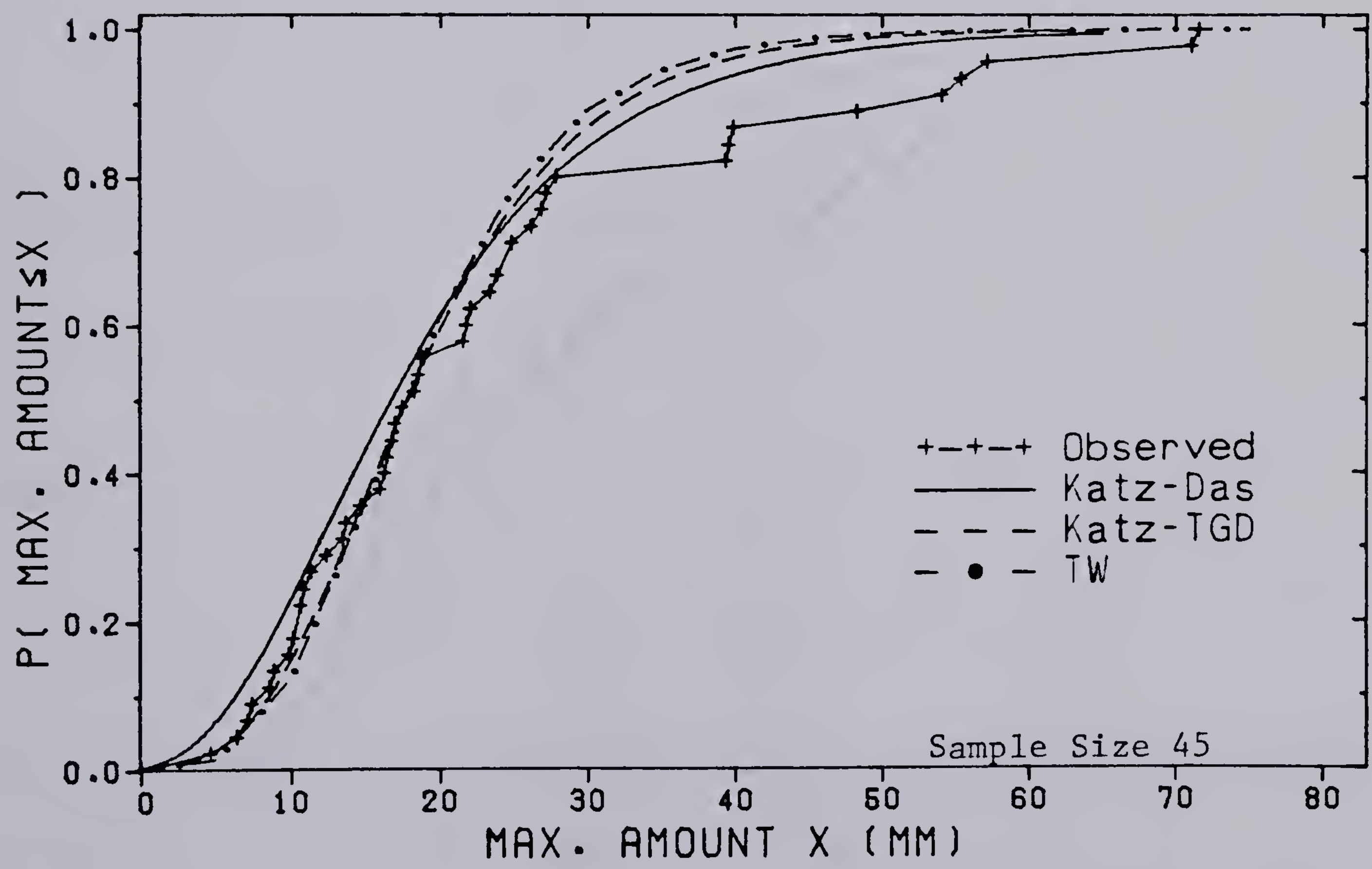


Figure 66. The theoretical distributions and the observed independent distribution for the maximum daily amount of precipitation in June at Medicine Hat.

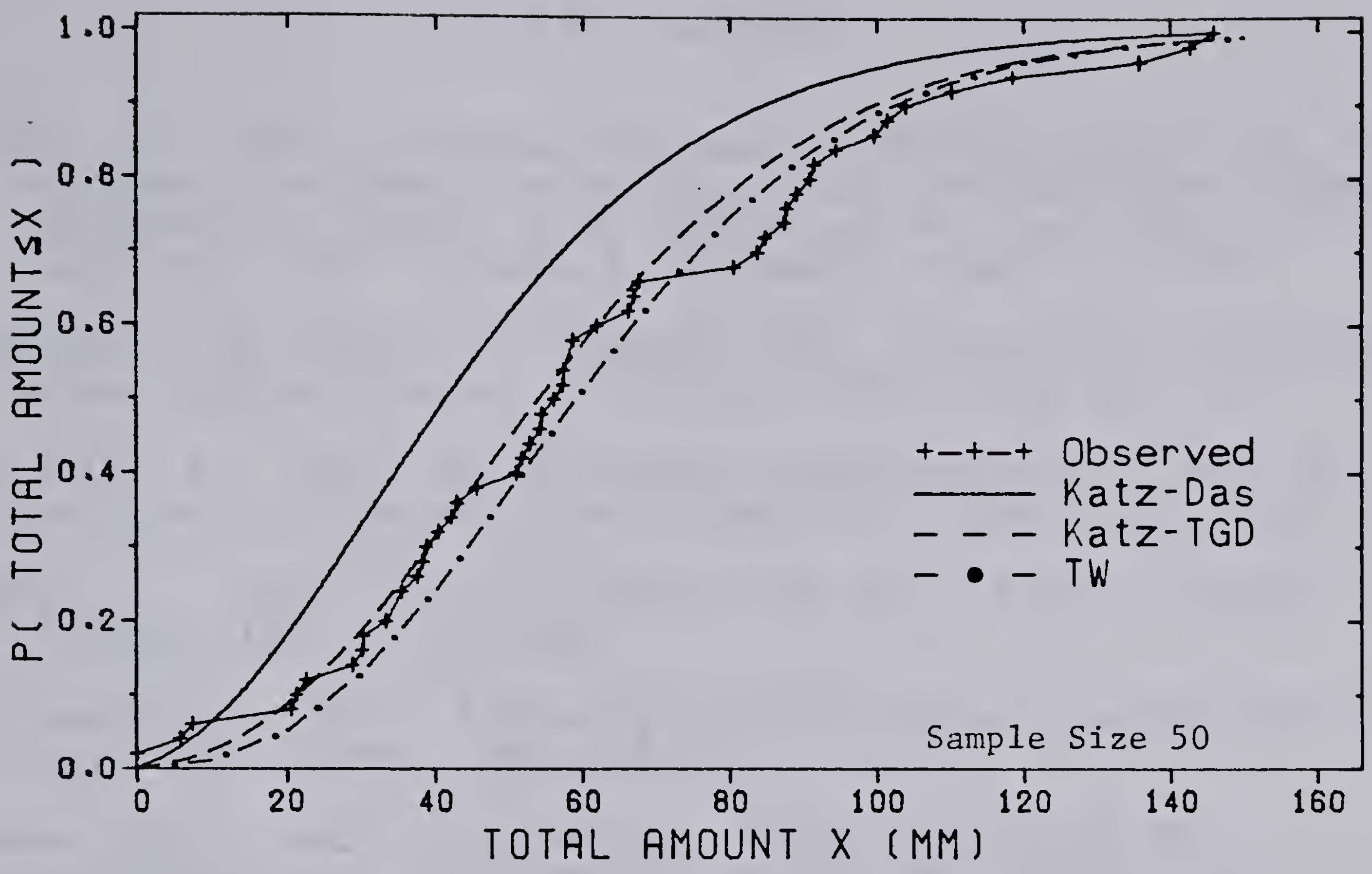


Figure 67. The theoretical distributions and the observed development distribution for the total amount of precipitation in June at Medicine Hat.

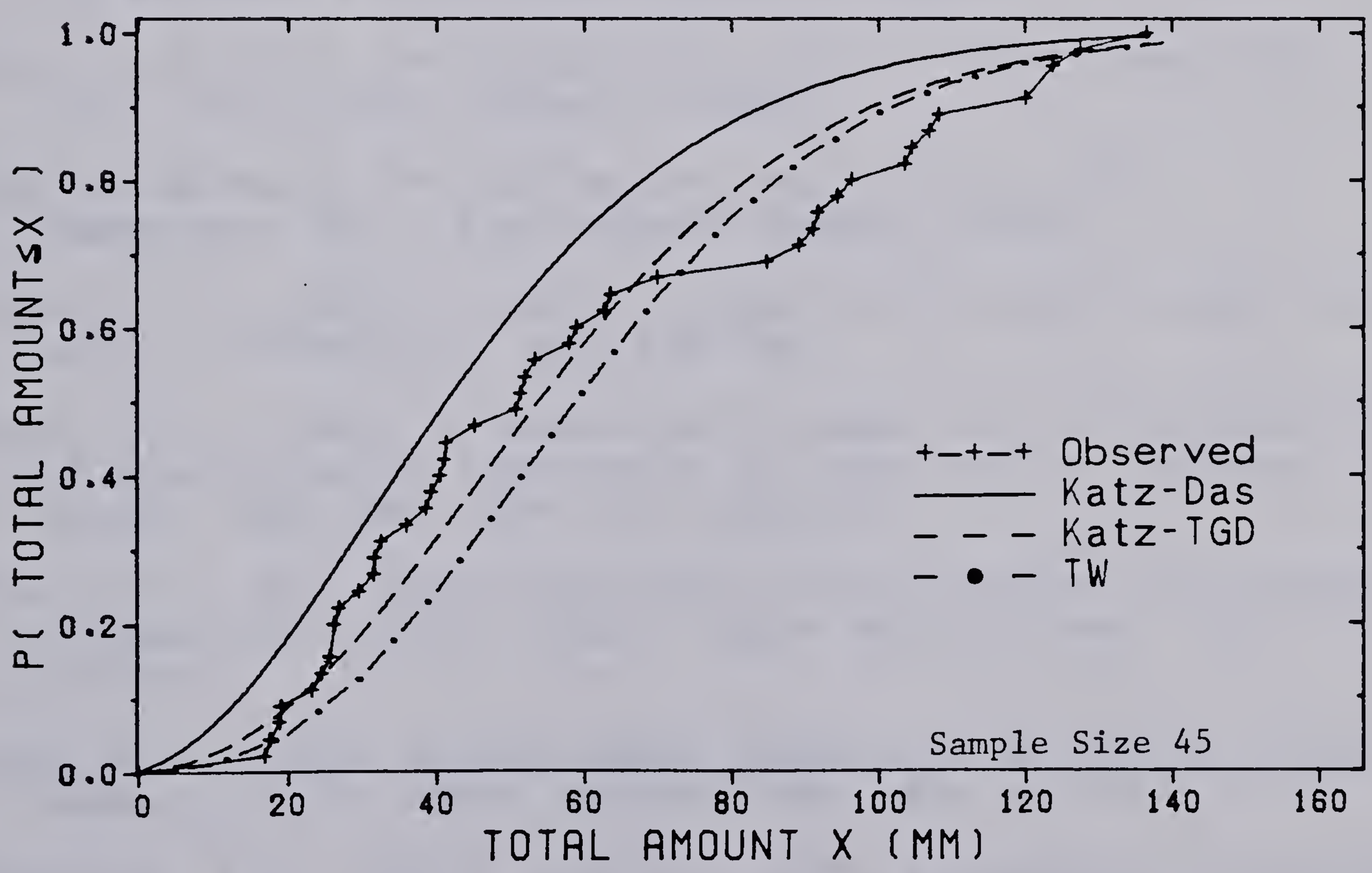


Figure 68. The theoretical distributions and the observed independent distribution for the total amount of precipitation in June at Medicine Hat.

Bibliography

Akaike, H., 1971: Information theory and an extension of the maximum likelihood principle. *Proc. Second Intern. Symp. Information Theory*, B.N. Petrov and F. Csaki eds., 2-8 September 1971, Budapest, Akademiai Kiado, 267-281.

Anderson, T.W., and L.A. Goodman,1957: Statistical inference about Markov chains. *Ann. Math. Stat.*, **28**, 89-110.

Bartlett, M., 1951: The frequency goodness of fit test for probability chains. *Proc. Camb. Phil. Soc.*, **47**, 86-95.

Besson, L., 1924: Sur la probabilité de la pluie. *Comptes Rendus*, **152**, 1743-1745.

Blackwell, D., 1953: Equivalent comparisons of experiments. *Ann. Math. Stat.*, **24**, 265-272.

Bowman, K.O., and L.R. Shenton, 1968: *Properties of Estimators for the Gamma Distribution*, Report CTC-1. Oak Ridge, Tenn., Union Carbide Corporation, Nuclear Division, 50pp.

Brooks, C.E.P., and N. Carruthers, 1953: *Handbook of Statistical Methods in Meteorology*. London, Her Majesty's Stationery Office, 412pp.

Canadian Normals, Precipitation, Vol. 2, 1973: Downsview, Ont., Environment Canada, 330pp.

Canadian Normals, Precipitation, Vol. 2-SI., 1975: Downsview, Ont., Environment Canada, 333pp.

Carder, A.C., 1962: Climatic trends in the Beaverlodge area. *Can. J. Plant Sci.*, **42**, 698-706.

Caskey, J.E., 1963: A Markov chain model for the probability of precipitation occurrence in intervals of various length. *Mon. Wea. Rev.*, **91**, 298-301.

Chin, E.H., 1977: Modelling daily precipitation occurrence process with Markov chain. *Water Resour. Res.*, **13**, 949-956.

Clark, R.R., 1979: A hydrologic reanalysis of the La Porte anomaly. *Bull. Amer. Meteor. Soc.*, **60**, 415-421.

Cleveland, W.S., and B. Kleiner, 1975: A graphical technique for enhancing scatterplots with moving averages. *Technometrics*, **17**, 447-454.

Climatic Normals, Vol. 2, 1968: Toronto, Meteorological

Branch, Canadian Dept. of Transport, 110pp.

Climatic Summaries for Selected Meteorological Stations in the Dominion of Canada, Vol. 1, 1947: Toronto, Meteorological Division, Canadian Dept. of Transport, 63pp.

Climatic Summaries for Selected Meteorological Stations in Canada, Addendum to Vol. 1, 1954: Toronto, Meteorological Division, Canadian Dept. of Transport, 29pp.

Climatological Station Data Catalogue, 1976: Downsview, Ont., Environment Canada, 152pp.

Cox, D.R., and H.D. Miller, 1965: *The Theory of Stochastic Processes.* New York, Wiley, 398pp.

Crutcher, H.L., 1975: A note on the possible misuse of the Kolmogorov-Smirnov test. *J. Appl. Meteor.*, 14, 1600-1603.

Das, S.C., 1955: The fitting of truncated type III curves to daily rainfall data. *Aust. J. Phys.*, 8, 298-304.

Farmer, E.E., and J.W. Homeyer, 1974: Probability of consecutive rainless days. *Water Resour. Bull.*, 10, 914-924.

Feller, W., 1957: *An Introduction to Probability Theory and its Applications.* Vol. 1, 2nd Ed., New York, John Wiley and Sons, 461pp.

Feyerherm, A.M., and L.D. Bark, 1964: *Probabilities of Sequences of Wet and Dry Days in Kansas.* Kans. Agr. Expt. Sta. Tech. Bull. 139, Manhattan, 55pp.

----, 1965: Statistical methods for persistent precipitation patterns. *J. Appl. Meteor.*, 4, 320-328.

----, 1967: Goodness of fit of a Markov chain model for sequences of wet and dry days. *J. Appl. Meteor.*, 6, 770-773.

Flueck, J.A., and P.W. Mielke Jr., 1975: Some bivariate distributions having meteorological applications. *Preprints Fourth Conf. on Probability and Statistics in Atmospheric Sciences*, 18-20 November 1975, Tallahassee, Amer. Meteor. Soc., 65-69.

Fox, D., and the staff of the Statistical Research Laboratory, 1976: *Elementary Statistics Using Midas*, 2nd Ed., University of Michigan, 300pp.

Gabriel, K.R., 1959: The distribution of the number of successes in a sequence of dependent trials. *Biometrika*, **46**, 454-460.

Gabriel, K.R., and J. Neumann, 1962: A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Roy. Meteor. Soc.*, **88**, 90-95.

Gates, P., and H. Tong, 1976: On Markov chain modeling to some weather data. *J. Appl. Meteor.*, **15**, 1145-1151.

Gerald, C.F., 1978: *Applied Numerical Analysis.* Addison Wesley, 518pp.

Good, I.J., 1955: The likelihood ratio test for Markov chains. *Biometrika*, **42**, 531-533.

Green, J., 1970: A generalized probability model for sequences of wet and dry days. *Mon. Wea. Rev.*, **98**, 238-241.

Greenwood, A.J., and D. Durand, 1960: Aids for fitting the gamma distribution. *Technometrics*, **2**, 55-65.

Haan, C.T., D.M. Allen, and J.O. Street, 1976: A Markov chain model of daily rainfall. *Water Resour. Res.*, **12**, 443-449.

Haan, C.T., 1977: *Statistical Methods in Hydrology.* Ames, Iowa, Iowa State University Press, 378pp.

Hannan, E.J., 1955: A test for singularities in Sydney rainfall. *Aust. J. Phys.*, **8**, 289-297.

Helgert, H.J., 1970: On sums of random variables defined on a two-state Markov chain. *J. Appl. Prob.*, **7**, 761-765.

Hoel, P.F., 1954: A test for Markov chains. *Biometrika*, **41**, 430-433.

Hopkins, J., and P. Robillard, 1964: Some statistics of daily rainfall occurrence for the Canadaian prairie provinces. *J. Appl. Meteor.*, **3**, 600-602.

IMSL Library, 1979: 7th Ed., Houston, Texas.

Ison, N.T., A.M. Feyerherm, and L.D. Bark, 1971: Wet period precipitation and the gamma distribution. *J. Appl. Meteor.*, **10**, 658-665.

Jorgensen, D.L., 1949: Persistency of rain and no rain periods during the winter at San Francisco. *Mon. Wea. Rev.*, **77**, 303-307.

Katz, R.W., 1974: Computing probabilities associated with the Markov chain model for precipitation. *J. Appl. Meteor.*, **13**, 953-954.

----, 1977a: Precipitation as a chain-dependent process. *J. Appl. Meteor.*, **16**, 671-676.

----, 1977b: Techniques for detecting dependence between meteorological variables. *Fifth Conf. on Probability and Statistics in Atmospheric Sciences*, 15-18 November 1977, Las Vegas, Amer. Meteor. Soc., 101-105.

----, 1977c: An application of chain-dependent processes to meteorology. *J. Appl. Prob.*, **14**, 598-603.

----, 1979a: Estimating the order of a markov chain: another look at the Tel Aviv rainfall data. *Preprints Sixth Conf. Probability and Statistics in Atmospheric Sciences*, 9-12 October 1979, Banff, Amer. Meteor. Soc., 217-221.

----, 1979b: *On some criteria for estimating the order of a Markov chain.* Presented at Annual Meeting of the American Statistical Association, 13-16 August 1979, Washinton, D.C., 18pp.

Kavvas, M.L., A.A. Aksit, and Y.K. Tulunay, 1977: A first order nonhomogeneous Markov chain for the daily rainfall occurrences in Ankara. In *Modeling Hydrologic Processes, Proc. of the Fort Collins Third Intern. Hydology Symp., on Theoretical and Applied Hydology*, 27-29 July 1977, Fort Collins, Colorado, Water Resources Publications, 44-59.

Klemeš, V., and A. Bulu, 1979: Limited confidence in confidence limits derived by operational stochastic hydrologic models. *J. Hydrol.*, **42**, 9-22.

Kendall, M.G., and A. Stuart, 1963: *The Advanced Theory of Statistics*. Vol. 1, 2nd Ed., Griffin, 433pp.

----, 1967: *The Advanced Theory of Statistics*. Vol.2, 2nd Ed., Griffin, 690pp.

Kullback, S., 1959: *Information Theory and Statistics*. Wiley, 395pp.

Lachapelle, P.A. 1977: *Modern Spectral Analysis of Alberta Climate*. MSc thesis, University of Alberta, 139pp.

Longley, R.W., 1953: The length of dry and wet periods. *Quart. J. Roy. Meteor. Soc.*, **79**, 520-527.

Lowry, W.P., and D. Guthrie, 1968: Markov chains of order

greater than one. *Mon. Wea. Rev.*, **96**, 798-801.

MANOBS, 1971: Manual of standard procedures for surface weather observing and reporting, Atmospheric Environment Service, Environment Canada, 307pp.

MANOBS, 1976: Manual of standard procedures for surface weather observing and reporting, Atmospheric Environment Service, Environment Canada, 439pp.

Meheriuk, W., 1972: *The Application of the Theory of Stochastic Processes to Precipitation at some Alberta Stations*. MSc Thesis, University of Alberta, 168pp.

Mielke, P.W., 1976: Simple iterative procedures for two-parameter gamma distribution maximum likelihood estimates. *J. Appl. Meteor.*,**15**, 181-183.

Potter, K.W., 1976: Evidence for nonstationarity as a physical explanation of the Hurst phenomenon. *Water Resour. Res.*, **12**, 1057-1052.

Richardson, C., 1977: *A Model of Stochastic Structure of Daily Precipitation over an Area*, Hydrology papers No. 91. Fort Collins, Colorado State University, 46pp.

Rodríguez, I., and V. Yevyevich, 1967: Sunspots and hydrologic time series. *Proc. of the Intern. Hydrology Symp.*, 6-8 September 1967, Fort Collins, Colorado, Colorado State University, 397-405.

Schickedanz, P.T., and G.F. Krause, 1970: A test for the scale parameters of two gamma distributions using the generalized likelihood ratio. *J. Appl. Meteor.*, **9**, 13-16.

Schwartz, G., 1978: Estimating the dimension of a model. *Ann. of Stat.*, **6**, 461-464.

Selvalingam, S., and M. Miura, 1978: Stochastic Modeling of monthly and daily rainfall sequences. *Water Resour. Bull.*, **14**, 1105-1120.

Skees, P.M., and L.R. Shenton, 1971: Comments on the statistical distribution of rainfall per period under various transformations. *Proc. Symp. Stat. Hydrol.*, 31 August-2 September 1971, Tucson, Arizona, USDA Misc. Publ. 1275, 172-196.

Thom, H.C., 1951: A frequency distribution for precipitation. Abstract, *Bull. Amer. Meteor. Soc.*, **32**, 397.

Thom, H.C., 1958: A note on the gamma distribution.

Mon. Wea. Rev., **86**, 117-122.

Thom, H.C., 1968: *Direct and Inverse Tables of the Gamma Distribution*, ESSA Technical Report, Environmental Data Service, Dept. of Commerce, Environmental Science Service Administration, Silver Spring, U.S.A., 30pp.

Todorovic, P., and D. Woolhiser, 1974: Stochastic models of daily rainfall. *Proc. Symp. on Stat. Hydrol.*, 31 August-2 September 1971, Tucson, Arizona, USDA Misc. Publ. 1275, 232-246.

----, 1975: A stochastic model of n-day precipitation. *J. Appl. Meteor.*, **14**, 17-24.

Todorovic, P., and V. Yevyevich, 1967: A particular stochastic process as applied to hydrology. *Proc. Intern. Hydrol. Symp.*, 6-8 September 1967, Fort Collins, Colorado, Colorado State University, 298-305.

Tong, H., 1975: Determination of the order of Markov chain by Akaike's information criterion. *J. Appl. Prob.*, **12**, 488-497.

Topil, A.G., 1963: Precipitation probability at Denver related to length of a period. *Mon. Wea. Rev.*, **91**, 293-297.

Tukey, J.W., 1977: *Exploratory Data Analysis.* Addison Wesley, 688pp.

Verschuren, J.P., 1968: *A Stochastic Analysis of Precipitation.* Ph. D. thesis, Colorado State University, Fort Collins, Colorado.

Weisner, C.J., 1970: *Hydrometeorology.* Chapman and Hall, London.

Weiss, L.L., 1944: *Preliminary Report on Duration of Stormy Periods at Selected Localities and Intervals Between Periods.* Research paper No. 3, U.S. Weather Bureau, Washington.

----, 1964: Sequences of wet or dry days described by a Markov chain probability model. *Mon. Wea. Rev.*, **92**, 169-176.

Wiser, E.H., 1965: Modified Markov probability models of sequences of precipitation events. *Mon. Wea. Rev.*, **93**, 511-516.

Wong, R., 1980: *GAM2*. Personal communication.

Woolhiser, D.A., E. Rovey, and P. Todorovic, 1973: Temporal

and spatial variation of parameters for the distribution of N-day precipitation. *Floods and Droughts, Preprints Second Intern. Symp. in Hydrol.*, 11-13 September 1972, Fort Collins, Colorado, Water Resources Publications, 605-614.

Woolhiser, D.A., and G.G. Pegram, 1979: Maximum likelihood estimation of Fourier coefficient to describe seasonal variation of parameters in stochastic daily precipitation models. *J. Appl. Meteor.*, 18, 34-42.

Yevyevich, V., 1972: *Structural Analysis of Hydrologic Time Series*, Hydrology papers No. 56. Fort Collins, Colorado, Colorado State University, 59pp.

Appendix A

The Markov chain is named after A.A. Markov who introduced the finite Markov chain in 1907 (Cox and Miller, 1965). An r-th order Markov chain is defined to be a sequence of discrete random variables Y_0, Y_1, ... with the property that the conditional distribution of Y_t depends on Y_{t-1}, Y_{t-2}..., Y_{t-r}, but not on Y_{t-r-1}, Y_{t-r-2}, ... Denote the S discrete states which the Y_t's assume by i, j, k=1, 2...S. For the first order or simple two-state Markov chain which was used in this study the r is 1 and S is 2.

In general, an r-th order chain may be reduced to a simple Markov chain by a redefinition of the state space (Cox and Miller, 1965). Consequently, this discussion will be limited to the case of a simple Markov chain, and the size of the state space is limited to 2.

The simple Markov chain is characterized by the property

$$Pr(Y_t=k|\ Y_{t-1}=j,\ Y_{t-2}=i,\ldots)=Pr(Y_t=k|\ Y_{t-1}=j).$$

The transition probability $Pr(Y_t=k|\ Y_{t-1}=j)$ is denoted $p_{ij}(t)$, i,j=1,2, and is the probability of an i-to-j transition at time t. The transition probabilities can be written in the form of a stochastic matrix

$$P(t) = \begin{bmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{bmatrix} \quad \text{for which}$$

$\sum_{j-1}^{2} p_{ij}(t)=1.$

Let $p_j(t)$ denote the probability that the state is in

state j at time t. Then $p(t)=(p_0(t), p_1(t))$ denotes the probability of each of the states being occupied at time t. Because

$$p_0(t)=p_0(t-1)\ p_{00} + p_1(t-1)\ p_{10} \text{ and}$$
$$p_1(t)=p_0(t-1)\ p_{01} + p_1(t-1)\ p_{11}$$

we have the result

$$p(t)=p(t-1)P,$$

so $p(t)=p(0)P^t$ where $p(0)$ is the initial probability distribution of the chain and P^t is the matrix of t step transition probabilities denoting Pr(chain is in state j at time t| chain is in state i at time 0).

If $p_{ij}(t)=p_{ij}(t+\tau)=p_{ij}$ for all τ, the Markov chain is said to be homogeneous.

A Markov chain may be further classified according to the classification of its states. A state k is classified according to the properties of the transition probabilties p_{jk}.

A state k is termed periodic if for any integer l>1 $p_{kk}(t)=0$, for t not an integral multiple of l. Moreover, if the return to a state k at some future time is a certain event the state is termed recurrent. If the mean recurrence time for a state k is finite the state is said to be positive recurrent.

The states of the Markov chain examined here were both aperiodic and positive recurrent.

An aperiodic, positive recurrent chain is termed ergodic. This means a unique limiting distribution

$$\bar{\pi} = (\pi_0, \pi_1)$$

exists and is given by,

$$\lim_{t\to\infty} P^t = \begin{pmatrix} \bar{\pi} \\ \bar{\pi} \end{pmatrix}.$$

For any p(0),

$$\lim_{t\to\infty} p(t) = \lim_{t\to\infty} p(0)P^t = p(0)\begin{pmatrix} \bar{\pi} \\ \bar{\pi} \end{pmatrix} = \bar{\pi}.$$

Recall that if the chain is initially in state k then p(0) has a 1 in the k-th position and the remaining p_j are 0. The limiting distribution $\bar{\pi}$ is termed the stationary distribution since if $p(0)=\bar{\pi}$ then $p(t)=\bar{\pi}$ for all t. For the two-state Markov chain used in this study

$$\pi_0 = \frac{p_{10}}{p_{01}+p_{10}} \quad \text{and} \quad \pi_1 = \frac{p_{01}}{p_{01}+p_{10}}$$

because $\bar{\pi} = \bar{\pi} P$ so $\bar{\pi}|I-P|=0$ with the constraint $\pi_0+\pi_1=1$.

For a long time $(t\to\infty)$ the proportion of time the chain spends in state k is just $\pi_k = 1/\bar{t}_k$ where $\bar{t}_k$ is the mean recurrence time for state k.

For a further introduction to Markov chains, reference should be made to Cox and Miller (1965) or Feller (1957).

Appendix B

The routines:

1. COUNT,
2. MDATUM,
3. MARKOV,
4. FOUR,
5. TW MARKOV CHAIN EXPONENTIAL MODEL,
6. KATZ DISTRIBUTION MODEL,
7. MAXP,
8. TOTP,
9. GAM,
10. FSG,
11. DERIV,
12. SIMPS,
13. GAM2,
14. ITS,

were used in this study. They appear in the following section.

```
C                    PROGRAM COUNT AUGUST 30,1979
C
C   THIS ROUTINE TABULATES THE FREQUENCY OF WET AND DRY DAY
C   SEQUENCES FOR A SELECTED PERIOD OF TIME FROM A CLIMATO-
C   LOGICAL RECOO ONMAGNETIC TAPE
C
C   LAST MOOIFIED 79 11 04
C
C   THIS VERSION RESTARTS THE SEQUENCE COUNTING,FROM
C   THE DAYS FOR WHICH A PRECIPITATION VALUE WAS
C   MISSING
C*********************************************************
C...I/O DEVICES..5=INITIAL VALUES AND PARAMETERS
C                6=OUTPUT MESSAGES
C                7=STATION OATA ON MAGNETIC TAPE
C                8=FINAL COUNTS
C                9=PLOTFILE
C                10=FILE WITH FOURIER COEFFICIENTS
C...VARIABLES....COUNT1-COUNT5=CONTAIN TOTAL NUMBER OF WET
C                              ANO ORY DAYS FOR EACH OAY
C                              OF THE YEAR
C                STRING=SEQUENCE OF O'S AND 1'S REPRESENTING
C                       DRY AND WET OAYS (A VECTOR IN THIS
C                       ROUTINE, PROBABLY MORE EFFICIENT TO
C                       MANIPULATE BITS)
C                PCPN=AMOUNT OF DAILY PRECIPITATION RECOROEO
C                OPTM=NUMBER OF DAYS IN YEAR PRIOR TO
C                     CURRENT MONTH
C                NDIM=NUMBER OF DAYS IN CURRENT MONTH
C                TNMOS=TOTAL NUMBER OF EACH MONTH RECOROED
C                STNIO=STATION IOENTIFICATION NUMBER
C                YR=THREE OIGIT YEAR
C                MO=TWO DIGIT MONTH
C                LEAPYR=VECTOR CONTAINING THREE DIGIT LEAP
C                       YEARS
C                FM=CHARACTER VARIABLE FOR MISSING OATA FLAG
C                FC=CHARACTER VARIABLE FOR PRECIPITATION
C                   OCCURREO BUT AMOUNT NOT RECOROEO
C                   FLAG
C                MSGO=NUMBER OF MISSING DAYS
C                FNM=MISSING DAY FLAG
C*********************************************************
C
C...SUBROUTINES CALLED INCLUDE:
C                MOATUM...FOR MISSING DATA VALUES
C                FOUR...CALCULATES FOURIER SERIES COEFFICIENTS
C                       AND PLOTS CUMULATIVE PERIODOGRAMS
C                MARKOV...CALCULATED AIC AND SBC
C                OUTPUT...AUXILLIARY ROUTINES FOR OTHER OUTPUT
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C...INITIALIZATION OF VALUES ANO INITIAL REAO STATEMENTS
C
      LOGICAL*1 LFMT(1) /'*'/
C...DIMENSION AND INITIALIZE APPROPRIATE ARRAYS AND COUNTERS
      INTEGER COUNT1(365,2), COUNT2(365,4), COUNT3(365,8),
     1        COUNT4(365,16), COUNT5(365,32), STRING(5),
     2        PCPN(31), DPTM(12), TNMOS(12), STNID, YR, FNM
      DIMENSION LEAPYR(25), NOIM(12)
      INTEGER*2 FLAG(31), FM, FC
      DATA STRING /5*O/, DPTM /O, 31, 59, 90, 120, 151, 181,
     1     212, 243, 273, 304, 334/, MSGO /O/, NOIM /31, 28,
     2     31, 30, 31, 30, 31, 31, 30, 31, 30, 31/,
     3     TNMOS /12*O/, FM /'M'/, FC /'C'/, LEAPYR /976,
     4     972, 968, 964, 960, 956, 952, 948, 944, 940, 936,
```

```
     5     932, 928, 924, 920, 916, 912, 908, 904, 900, 896,
     6     892, 888, 884, 880/
C...EARLIEST LEAP YEAR IS 1880
      DO 9 I=1,365
      DO 1 J=1,2
    1 COUNT1(I,J)=0
      DO 2 J=1,4
    2 COUNT2(I,J)=0
      DO 3 J=1,8
    3 COUNT3(I,J)=0
      DO 4 J=1,16
    4 COUNT4(I,J)=0
      DO 5 J=1,32
    5 COUNT5(I,J)=0
    9 CONTINUE
C...PROMPT FOR STATION I. D., INITIAL DATE, AND  FINAL DATE
C...INITIAL DAY SHOULD NOT BE LATER THAN THE 23 RD DAY OF
C   THE MONTH SO THAT IF THE INITIAL MONTH IS FEBRUARY
C   THE STRING INITIALIZATION WILL BE COMPLETED PRIOR
C   TO ANOTHER READ STATEMENT BEING NECESSARY
      WRITE (6,10)
   10 FORMAT ('0', 'INSERT STATION I.D. AND INITIAL, 3 DIG',
     1        'IT YEAR, TWO DIGIT MONTH, AND 2 DIGIT DAY'/1X,
     2        'THE INITIAL DAY SHOULD NOT BE LATER THAN THE',
     3        ' 23 RD DAY OF THE MONTH, FINAL 3 DIGIT YEAR')
C...INPUT INITIAL STATION I. D. AND DATE  FINAL YEAR
      READ (5,LFMT) ISTNID, IYR, IMO, IDAY, IFYR
      WRITE (6,20) ISTNID, IYR, IMO, IDAY, IFYR
   20 FORMAT ('0', 'INITIAL STATION I.D.', I9, ' INITIAL ',
     1        'YEAR', I5, ' INITIAL MONTH', I4, ' INITIAL ',
     2        'DAY',I4/' FINAL YEAR',I4)
C...DATA SHOULD BE CHECKED FOR COMPLETENESS AND CONTINUITY
C   PRIOR TO APPLICATION OF COUNT PROGRAM
C...INPUT INITIAL MONTH OF CLIMATE RECORD
   30 READ (7,40) STNID, YR, MO, (PCPN(I),FLAG(I),I=1,31)
   40 FORMAT (I7, I3, I2, 3X, 31(I6,A1))
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C...POSITION TAPE TO CORRECT INITIAL MONTH
C
C...CHECK FOR CORRECT STATION, IF INCORRECT THE TAPE/FILE IS
C   WRONG, TERMINATE PROGRAM
      IF (STNID .EQ. ISTNID) GO TO 60
      WRITE (6,50)
   50 FORMAT ('0', 'STATION IDENTIFICATION INCORRECT',
     1        ' PROGRAM TERMINATED')
      STOP
C...CHECK FOR CORRECT STARTING DATE AND POSITION TAPE/FILE.
C   IF NECESSARY
   60 IF (YR .LE. IYR) GO TO 80
      WRITE (6,70)
   70 FORMAT ('0', 'FIRST YEAR READ LATER THAN INITIAL YEAR'
     1        )
      STOP
   80 IF (MO .LE. IMO .OR. YR .LT. IYR) GO TO 100
      WRITE (6,90)
   90 FORMAT ('0', 'FIRST MONTH READ LATER THAN INITIAL ',
     1        'MONTH')
      STOP
  100 IF (MO .EQ. IMO) GO TO 110
      GO TO 30
  110 IF (YR .EQ. IYR) GO TO 120
C
C  NOTE THAT IF A MONTH OF RECORD IS MISSINGDURING THE
C  NEXT 11 MONTHS THE ROUTINE WILL STOP WITH A FIRST
C  YEAR READ LATER THAN INITIAL YEAR MESSAGE, EVEN IF
```

```
C  IT WAS NOT THE FIRST YEAR READ, NECESSARY TO SKIP
C  RECORDS PRIOR APPLICATION OF ROUTINE
C
C  SKIP IS U.OF A. SYSTEM SUBROUTINE FOR POSITIONING TAPE
C
      CALL SKIP(0, 11, 7, &570, &570, &550)
      GO TO 30
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C...SHOULD ENSURE THAT THERE ARE 5 SEQUENTIAL DAYS
C   OF NON MISSING OATA IN THE FIRST MONTH,
C   OTHERWISE MOATUM WILL CHANGE MONTHS
C
C...DEALS WITH MISSING INITIAL DAY
C
  120 CALL MOATUM(MSGO,YR,MO,IOAY,FLAG,FNM,NOIM)
C  IF THERE WAS NOT FIVE CONSECUTIVE OAYS OF RECOROS
C  OR THERE IS A MISSING OAY OURING THE LAST FIVE OAYS
C  OF THE MONTH, REAO A NEW RECORO AT 334
C
      IF(FNM.EQ.1) GO TO 334
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C...INITIALIZATION OF SEQUENCE
C
C...I IS THE DAY (1 TO 365) OF THE YEAR
      L=IDAY
  180 IDAY=L
C...RESET FLAG
      FNM=5
      I = DPTM(MO) + IDAY
C...IF MEASURED PRECIPITATION WAS REPORTED FOR THE DAY OR
C   IF PRECIPITATION OCCURRED BUT THE AMOUNT IS UNKNOWN SET
C   THE STRING INOICATOR TO 1, O OTHERWISE.  TABULATE THE
C   OCCURRENCE OF A WET (COUNT1(I,2)) OR A DRY (COUNT1(I,1))
C   DAY
      IF (PCPN(IDAY) .GT. O .OR. FLAG(IOAY) .EQ. FC)
     1   GO TO 190
      STRING(5) = O
      COUNT1(I,1) = 1 + COUNT1(I,1)
      GO TO 200
  190 STRING(5) = 1
      COUNT1(I,2) = 1 + COUNT1(I,2)
C...INCREMENT THE OAY OF THE YEAR
  200 I = I + 1
      IF (PCPN(IOAY + 1) .GT. O .OR. FLAG(IDAY + 1) .EQ. FC)
     1   GO TO 210
      STRING(4) = O
      COUNT1(I,1) = 1 + COUNT1(I,1)
      GO TO 220
  210 COUNT1(I,2) = 1 + COUNT1(I,2)
      STRING(4) = 1
C...CALCULATE STORAGE LOCATION FOR TABULATION OF THE 2
C   SEQUENCE
  220 K2 = STRING(5) * 2 + STRING(4) + 1
      COUNT2(I,K2) = 1 + COUNT2(I,K2)
C...INCREMENT THE DAY OF THE YEAR
      I = I + 1
      IF (PCPN(IDAY + 2) .GT. O .OR. FLAG(IDAY + 2) .EQ. FC)
     1   GO TO 230
C...TABULATE WET OR ORY DAY
      COUNT1(I,1) = 1 + COUNT1(I,1)
      STRING(3) = O
      GO TO 240
  230 COUNT1(I,2) = 1 + COUNT1(I,2)
      STRING(3) = 1
C...CALCULATE STORAGE LOCATION FOR 2 ANO 3 DAY SEQUENCE
```

```
C   COUNT AND TABULATE THE SEQUENCES
  240 K2 = STRING(4) * 2 + STRING(3) + 1
      K3 = STRING(5) * 4 + K2
      COUNT2(I,K2) = 1 + COUNT2(I,K2)
      COUNT3(I,K3) = 1 + COUNT3(I,K3)
C...INCREMENT DAY OF THE YEAR
      I = I + 1
      IF (PCPN(IDAY + 3) .GT. O .OR. FLAG(IDAY + 3) .EQ. FC)
     1   GO TO 250
C...TABULATE THE WET OR DRY DAY
      COUNT1(I,1) = 1 + COUNT1(I,1)
      STRING(2) = O
      GO TO 260
  250 COUNT1(I,2) = 1 + COUNT1(I,2)
      STRING(2) = 1
C...CALCULATE STORAGE LOCATION FOR 2 TO 4 DAY SEQUENCE COUNT
C   TABULATE THE SEQUENCES AND SHIFT STRING TO EARLIER DAY
  260 K2 = STRING(3) * 2 + STRING(2) + 1
      K3 = STRING(4) * 4 + K2
      K4 = STRING(5) * 8 + K3
      COUNT2(I,K2) = 1 + COUNT2(I,K2)
      COUNT3(I,K3) = 1 + COUNT3(I,K3)
      COUNT4(I,K4) = 1 + COUNT4(I,K4)
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C...TABULATION OF SEQUENCES
C
C...AT THIS POINT THE FULL 5 SEQUENCE OF DAYS HAS BEEN
C   INITIALIZED SO USE A LOOP TO COMPLETE THE TABULATION
C   FOR THE INITIAL MONTH
      J = IDAY + 4
  270 K = NOIM(MO)
C...LY, A FLAG TO INDICATE THE OCCURRENCE OF 29 DAYS IN FEB
C...CHECK FOR THE OCCURRENCE OF 29 DAYS IN FEB,
C   IF YES SET LY=1,
C   THE 29TH WILL BE USED IN THE SEQUENCES BUT NO TABULATION
C   WILL BE MADE FOR THE 29TH DAY OF FEB.
      LY = O
      IF (MO .NE. 2) GO TO 290
      DO 280 IJ = 1, 25
        IF (LEAPYR(IJ) .NE. YR) GO TO 280
        LY = 1
        K = 29
  280 CONTINUE
  290 DO 320 L = J, K
C...SET DAY OF THE YEAR
        I = DPTM(MO) + L
C...DETERMINE IF THE PRECIPITATION VALUE IS MISSING, IF SO
C   RE-INITIALIZE SEQUENCE
        IF (FLAG(L) .EQ. FM) CALL MDATUM(MSGD, YR, MO, L,
     1         FLAG,FNM,NOIM)
C  IF HAVE 5 CONSECUTIVE REPORTS IN MONTH AFTER
C  MISSING DAY REINITIALIZE SEQUENCE IN CURRENT MONTH
      IF(FNM.EQ.O) GO TO 180
C IF MISSING VALUE OCCURRED IN LAST 5 DAYS OF MONTH
C NEED TO READ NEXT MONTH AND INCREMENT APPROPRIATE COUNTERS
      IF(FNM.EQ.1) GO TO 334
C...DETERMINE IF THE CURRENT DAY IS WET OR DRY AND SET
C   STRING(1) ACCORDINGLY
        IF (PCPN(L) .GT. O .OR. FLAG(L) .EQ. FC)
     1     GO TO 300
        STRING(1) = O
        GO TO 310
  300   STRING(1) = 1
C...CALCULATE STORAGE LOCATIONS FOR THE 1-5 SEQUENCES
  310   K1 = STRING(1) + 1
```

```
         K2 = STRING(2) * 2 + K1
         K3 = STRING(3) * 4 + K2
         K4 = STRING(4) * 8 + K3
         K5 = STRING(5) * 16 + K4
C...SHIFT STRING 1 DAY
         STRING(5) = STRING(4)
         STRING(4) = STRING(3)
         STRING(3) = STRING(2)
         STRING(2) = STRING(1)
C...CHECK FOR THE 29TH DAY OF FEB., DO NOT TABULATE
C   SEQUENCES FOR THIS DAY
         IF (LY .EQ. 1 .AND. L .EQ. 29) GO TO 330
         COUNT1(I,K1) = 1 + COUNT1(I,K1)
         COUNT2(I,K2) = 1 + COUNT2(I,K2)
         COUNT3(I,K3) = 1 + COUNT3(I,K3)
         COUNT4(I,K4) = 1 + COUNT4(I,K4)
  320 COUNT5(I,K5) = 1 + COUNT5(I,K5)
C...TABULATE THE NUMBER OF EACH MONTH IN THE RECORD
C...NOTE..MONTHS FLAGGED AS HAVING A MISSING
C         DAY IN THE LAST FIVE OF THE MONTH
C         WILL NOT BE COUNTED
  330 TNMOS(MO) = 1 + TNMOS(MO)
C...AT THIS POINT, IN THE FIRST PASS,THE INITIAL MONTH OF
C   THE RECORD HAS BEEN TABULATED, READ THE NEXT MONTHS
C   RECORD AND CONTINUE THE TABULATION
  334 READ (7,40,END=340) STNID, YR, MO, (PCPN(I),FLAG(I),I=
     11,31)
C
C CHECK TO SEE IF FINISHED, AND IF NOT THAT THE NEXT
C MONTH OF RECORD FOLLOWS THE CURRENT MONTH IN THE YEAR
C
      IF(YR.EQ.IFYR) GO TO 340
      IMO=IMO+1
      IF(IMO.GT.12) GO TO 339
  335 IDAY=1
      IF(IMO.NE.MO) GO TO 336
      J = 1
      IF(FNM.EQ.1) GO TO 120
      GO TO 270
  336 WRITE(6,337) IYR,IMO
  337 FORMAT('0','THE MONTH ',I3,I2,' IS MISSING')
      IMO=IMO+1
      FNM=1
      IF(IMO.GT.12) GO TO 339
      GO TO 335
  339 IMO=1
      IYR=IYR+1
      GO TO 335
C+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C...OUTPUT THE TOTAL NUMBER OF MISSING DAYS AND THE
C   TOTAL NUMBER OF EACH MONTH WITH RECORDS
C
  340 WRITE (6,350) MSGD, (TNMOS(I),I=1,12)
  350 FORMAT ('0', 'THE TOTAL NUMBER OF MISSING DAYS',
     1         I6/'0', 'THE TOTAL NUMBER OF EACH ',
     2        'MONTH OBSERVED'/'0', 12I5)
C...CALL SUBROUTINE MARKOV TO CALCULATE MARKOV
C   CHAIN ORDER, INSERT NULL ROUTINE IF MARKOV
C   CHAIN ORDER IS NOT REQUIRED
      CALL MARKOV(COUNT1,COUNT2,COUNT3,COUNT4,COUNT5)
C...CALL SUBROUTINE FOUR TO FIT A FOURIER SERIES
C   TO THE DAILY PROBABILITY ESTIMATES IN COUNT1-5
C   INSERT A NULL ROUTINE IF ESTIMATES NOT REQUIRED
C
      CALL FOUR(COUNT1,COUNT2,COUNT3,COUNT4,COUNT5)
```

```
C...OUTPUT THE SEQUENCE TOTALS OR DO OTHER TESTS
C
      CALL OUTPUT(COUNT1,COUNT2,COUNT3,COUNT4,COUNT5)
      STOP 20
C++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
C...ERROR MESSAGES
C
C...ERROR MESSAGES FOR READ PROBLEMS WHEN TAPE OR FILE IS
C   BEING POSITIONED
  550 WRITE (6,560)
  560 FORMAT ('0', 'TAPE OR FILE DEVICE INCORRECT IN SKIP ',
     1       'ROUTINE')
      STOP
  570 WRITE (6,580)
  580 FORMAT ('0', 'END OF FILE OR TAPE REACHED WHEN ',
     1       'SEARCHING FOR CORRECT INITIAL DATE')
      STOP
      END
```

```
      SUBROUTINE MDATUM(MSGD, YR, MO, MDAY, FLAG, FNM, NDIM)
C
C...SUBROUTINE MDATUM NOTES THE MISSING DAYS
C   OF THE RECORD AND SEARCHES FOR A SERIES OF
C   FIVE CONSECUTIVE DAYS WITH NON-MISSING
C   PRECIPITATION VALUES, IT TABULATES THE TOTAL
C   NUMBER DF MISSING DAYS IN THE RECORD, NOT
C   CDUNTING THDSE DAYS MISSING DURING THE LAST
C   FIVE DAYS OF ANY MONTH
C
C   VARIABLES   NDIM=NUMBER OF DAYS IN MONTH
C               FNM=UTILITY FLAG
C               FM=VECTOR DF FLAGS FOR DAILY VALUES
C               MDAY=MISSING DAY OF MONTH
C   INITIALIZATIDN
C
      DIMENSIDN NDIM(12)
      INTEGER FNM, YR
      INTEGER*2 FLAG(31), FM
      DATA FM /'M'/
      FNM = 0
C CHECK TO SEE IF MDAY IS MISSING, IF YES FLAG IT AS
C MISSING AND INCREMENT DAY, IF DURING LAST FIVE
C DAYS IN MONTH DUTPUT MESSAGE AND SET APPROPRIATE FLAGS
C
   30 IF (FLAG(MDAY) .NE. FM) GO TD 50
      MSGD=MSGD+1
      WRITE (6,10) MDAY, YR, MO
   10 FORMAT ('0', 'THE ', I2, ' DAY DF ', I3, I2,
     1          ' IS MISSING')
      MDAY = MDAY + 1
      IF (MDAY .LE. NDIM(MD) - 4) GO TD 30
      WRITE (6,40) YR, MD
   40 FORMAT ('0', 'THE MDNTH ', I3, I2,' HAS A MISSING',
     1 ' VALUE DURING THE LAST FIVE DAYS OF THE',
     2' MONTH')
      FNM = 1
      GO TO 70
C
C  CHECK TO SEE IF NEXT 5 DAYS OF RECORD AVAILABLE
C
   50 DO 60 J = 1, 4
         K = J
         IF (FLAG(MDAY + J) .EQ. FM) GD TD 80
   60 CONTINUE
   70 RETURN
   80 MDAY = MDAY + K
      GO TD 30
      END
```

```
      SUBROUTINE MARKOV(COUNT1,COUNT2,COUNT3,COUNT4,COUNT5)
C
C...PROGRAM CREATED BY K. JOHNSTONE
C...LAST MODIFIED 80 04 30
C...PROGRAM CALCULATES THE AIC AND SBC CRITERION
C...REF. GATES AND TONG, 1976; KATZ, 1979A)
C
C
C...I/O  6=TABULATED COUNTS AND A. I. C. INFORMATION
C           ESTIMATES
C        5=DATES FOR CHAIN ESTIMATION
C
C...INITITALIZATION AND DIMENSIONING
C
      IMPLICIT REAL*8 (A-H,O-Z)
      LOGICAL*1 LFMT(1)/'*'/
      INTEGER COUNT1(365,2),COUNT2(365,4),COUNT3(365,8),COUNT4(365,16),
     1COUNT5(365,32), DATES(12,2)
      DIMENSION R(5), SB(5)
      COMMON /BLOCK1/NPRD,DATES
C
C...READ THE NUMBER OF PERIODS TO BE EXAMINED
C
      WRITE(6,100)
 100  FORMAT('OINPUT NUMBER OF PERIODS TO BE EXAMINED FOR',
     &' MARKOV CHAIN ORDER')
      READ(5,LFMT) NPRD
C
C...INPUT THE DATES FOR THE BEGINNING AND END OF THE PERIODS
C   ONE SET OF PERIOD VALUES PER LINE
C
      WRITE(6,101)
 101  FORMAT('OINPUT BEGINNING AND END DAY FOR EACH PERIOD',
     &' ONE SET OF VALUES PER LINE')
      DO 40 I=1,NPRD
 40   READ(5,LFMT) DATES(I,1),DATES(I,2)
C
C...EVALUATE AIC AND SBC ESTIMATES FOR EACH PERIOD
C
      DO 50 INDEX=1,NPRD
C
C...INITIALIZE VALUES TO ZERO
C
      D=0.D0
      W=0.D0
      WW=0.D0
      WD=0.D0
      DW=0.D0
      DD=0.D0
      WWW=0.D0
      WWD=0.D0
      DWW=0.D0
      DWD=0.D0
      WDW=0.D0
      WDD=0.D0
      DDW=0.D0
      DDD=0.D0
      WWWW=0.D0
      WWWD=0.D0
      DWWW=0.D0
      DWWD=0.D0
      WDWW=0.D0
      WDWD=0.D0
```

```
      DDWW=0.D0
      DDWD=0.D0
      WWDW=0.D0
      WWDD=0.D0
      DWDW=0.D0
      DWDD=0.D0
      WDDW=0.D0
      WDDD=0.D0
      DDDW=0.D0
      DDDD=0.D0
      WWWWW=0.D0
      WWWWD=0.D0
      WDWWW=0.D0
      WDWWD=0.D0
      WWDWW=0.D0
      WWDWD=0.D0
      WDDWW=0.D0
      WDDWD=0.D0
      WWWDW=0.D0
      WWWDD=0.D0
      WDWDW=0.D0
      WDWDD=0.D0
      WWDDW=0.D0
      WWDDD=0.D0
      WDDDW=0.D0
      WDDDD=0.D0
      DWWWW=0.D0
      DWWWD=0.D0
      DDWWW=0.D0
      DDWWD=0.D0
      DWDWW=0.D0
      DWDWD=0.D0
      DDDWW=0.D0
      DDDWD=0.D0
      DWWDW=0.D0
      DWWDD=0.D0
      DDWDW=0.D0
      DDWDD=0.D0
      DWDDW=0.D0
      DWDDD=0.D0
      DDDDW=0.D0
      DDDDD=0.D0
C
C...BEGIN TABULATIONS FOR THE PERIOD REQUIRED;
C   THE PERIOD REQUIRED IS OBTAINED BY SETTING THE
C   LIMITS OF THE DO TO THE DAYS REQUESTED
C
      JB=DATES(INDEX,1)
      JE=DATES(INDEX,2)
      DO 10 I=JB ,JE
      D=D+DFLOAT(COUNT1(I,1))
      W=W+DFLOAT(COUNT1(I,2))
      WW=WW+DFLOAT(COUNT2(I,4))
      WD=WD+DFLOAT(COUNT2(I,3))
      DW=DW+DFLOAT(COUNT2(I,2))
      DD=DD+DFLOAT(COUNT2(I,1))
      WWW=WWW+DFLOAT(COUNT3(I,8))
      WWD=WWD+DFLOAT(COUNT3(I,7))
      WDW=WDW+DFLOAT(COUNT3(I,6))
      WDD=WDD+DFLOAT(COUNT3(I,5))
      DWW=DWW+DFLOAT(COUNT3(I,4))
      DWD=DWD+DFLOAT(COUNT3(I,3))
      DDW=DDW+DFLOAT(COUNT3(I,2))
      DDD=DDD+DFLOAT(COUNT3(I,1))
      WWWW=WWWW+DFLOAT(COUNT4(I,16))
      WWWD=WWWD+DFLOAT(COUNT4(I,15))
```

```
      WWDW=WWDW+DFLOAT(COUNT4(I,14))
      WWDD=WWDD+DFLOAT(COUNT4(I,13))
      WDWW=WDWW+DFLOAT(COUNT4(I,12))
      WDWD=WDWD+DFLOAT(COUNT4(I,11))
      WDDW=WDDW+DFLOAT(COUNT4(I,10))
      WDDD=WDDD+DFLOAT(COUNT4(I,9))
      DWWW=DWWW+DFLOAT(COUNT4(I,8))
      DWWD=DWWD+DFLOAT(COUNT4(I,7))
      DWDW=DWDW+DFLOAT(COUNT4(I,6))
      DWDD=DWDD+DFLOAT(COUNT4(I,5))
      DDWW=DDWW+DFLOAT(COUNT4(I,4))
      DDWD=DDWD+DFLOAT(COUNT4(I,3))
      DDDW=DDDW+DFLOAT(COUNT4(I,2))
      DDDD=DDDD+DFLOAT(COUNT4(I,1))
      WWWWW=WWWWW+DFLOAT(COUNT5(I,32))
      WWWWD=WWWWD+DFLOAT(COUNT5(I,31))
      WWWDW=WWWDW+DFLOAT(COUNT5(I,30))
      WWWDD=WWWDD+DFLOAT(COUNT5(I,29))
      WWDWW=WWDWW+DFLOAT(COUNT5(I,28))
      WWDWD=WWDWD+DFLOAT(COUNT5(I,27))
      WWDDW=WWDDW+DFLOAT(COUNT5(I,26))
      WWDDD=WWDDD+DFLOAT(COUNT5(I,25))
      WDWWW=WDWWW+DFLOAT(COUNT5(I,24))
      WDWWD=WDWWD+DFLOAT(COUNT5(I,23))
      WDWDW=WDWDW+DFLOAT(COUNT5(I,22))
      WDWDD=WDWDD+DFLOAT(COUNT5(I,21))
      WDDWW=WDDWW+DFLOAT(COUNT5(I,20))
      WDDWD=WDDWD+DFLOAT(COUNT5(I,19))
      WDDDW=WDDDW+DFLOAT(COUNT5(I,18))
      WDDDD=WDDDD+DFLOAT(COUNT5(I,17))
      DWWWW=DWWWW+DFLOAT(COUNT5(I,16))
      DWWWD=DWWWD+DFLOAT(COUNT5(I,15))
      DWWDW=DWWDW+DFLOAT(COUNT5(I,14))
      DWWDD=DWWDD+DFLOAT(COUNT5(I,13))
      DWDWW=DWDWW+DFLOAT(COUNT5(I,12))
      DWDWD=DWDWD+DFLOAT(COUNT5(I,11))
      DWDDW=DWDDW+DFLOAT(COUNT5(I,10))
      DWDDD=DWDDD+DFLOAT(COUNT5(I,9))
      DDWWW=DDWWW+DFLOAT(COUNT5(I,8))
      DDWWD=DDWWD+DFLOAT(COUNT5(I,7))
      DDWDW=DDWDW+DFLOAT(COUNT5(I,6))
      DDWDD=DDWDD+DFLOAT(COUNT5(I,5))
      DDDWW=DDDWW+DFLOAT(COUNT5(I,4))
      DDDWD=DDDWD+DFLOAT(COUNT5(I,3))
      DDDDW=DDDDW+DFLOAT(COUNT5(I,2))
   10 DDDDD=DDDDD+DFLOAT(COUNT5(I,1))
C
C...DETERMINE TOTALS
C
         T11=(D+W)
         T21=(DD+DW)
         T23=(WD+WW)
         T31=(DDD+DDW)
         T33=(DWD+DWW)
         T35=(WDD+WDW)
         T37=(WWD+WWW)
         T41=(DDDD+DDDW)
         T43=(DDWD+DDWW)
         T45=(DWDD+DWDW)
         T47=(DWWD+DWWW)
         T49=(WDDD+WDDW)
         T411=(WDWD+WDWW)
         T413=(WWDD+WWDW)
         T415=(WWWD+WWWW)
         T51=(DDDDD+DDDDW)
         T53=(DDDWD+DDDWW)
```

```
      T55=(DDWDD+DDWDW)
      T57=(DDWWD+DDWWW)
      T59=(DWDDD+DWDDW)
      T511=(DWDWD+DWDWW)
      T513=(DWWDD+DWWDW)
      T515=(DWWWD+DWWWW)
      T517=(WDDDD+WDDDW)
      T519=(WDDWD+WDDWW)
      T521=(WDWDD+WDWDW)
      T523=(WDWWD+WDWWW)
      T525=(WWDDD+WWDDW)
      T527=(WWDWD+WWDWW)
      T529=(WWWDD+WWWDW)
      T531=(WWWWD+WWWWW)
C
C...BEGIN CALCULATIDN DF THE MAXIMUM LIKELIHDDD RATIO
C   TEST STATISTICS FOR TESTING THE NULL HYPOTHESIS THAT
C   THE CHAIN IS DF ORDER K < R
C...REF(TONG, 1975; GATES AND TDNG, 1976; HDEL, 1954;
C       GOOD, 1955)
C
C...RATIO TEST STATISTIC FOR HYPOTHESIS THAT CHAIN IS
C   OF ORDER 3 < 4
C
C...CALCULATION DONE IN 4 SECTIONS TD ELIMINATE SUBTRACTIONS
C   AND TO KEEP NUMBER OF CONTINUATIDN CARDS LESS THAN 19
C
      X1=DDDDD*DLDG(DDDDD/T51)
     1+DDDDW*DLDG(DDDDW/T51)
     2+DDDWD*DLDG(DDDWD/T53)
     3+DDDWW*DLDG(DDDWW/T53)
     4+DDWDD*DLOG(DDWDD/T55)
     5+DDWDW*DLDG(DDWDW/T55)
     6+DDWWD*DLDG(DDWWD/T57)
     7+DDWWW*DLOG(DDWWW/T57)
     8+DWDDD*DLDG(DWDDD/T59)
     9+DWDDW*DLDG(DWDDW/T59)
     &+DWDWD*DLDG(DWDWD/T511)
     1+DWDWW*DLDG(DWDWW/T511)
     2+DWWDD*DLOG(DWWDD/T513)
     2+DWWDW*DLDG(DWWDW/T513)
     3+DWWWD*DLDG(DWWWD/T515)
     4+DWWWW*DLDG(DWWWW/T515)
     5+WDDDD*DLOG(WDDDD/T517)
     6+WDDDW*DLOG(WDDDW/T517)
     7+WDDWD*DLOG(WDDWD/T519)
     8+WDDWW*DLOG(WDDWW/T519)
C
      X3=DDDDD*                 DLDG(DDDD/T41)
     1+DDDDW*                   DLOG(DDDW/T41)
     2+DDDWD*                   DLOG(DDWD/T43)
     3+DDDWW*                   DLOG(DDWW/T43)
     4+DDWDD*                   DLOG(DWDD/T45)
     5+DDWDW*                   DLDG(DWDW/T45)
     6+DDWWD*                   DLOG(DWWD/T47)
     7+DDWWW*                   DLOG(DWWW/T47)
     8+DWDDD*                   DLOG(WDDD/T49)
     9+DWDDW*                   DLOG(WDDW/T49)
     &+DWDWD*                   DLDG(WDWD/T411)
     1+DWDWW*                   DLOG(WDWW/T411)
     2+DWWDD*                   DLOG(WWDD/T413)
     2+DWWDW*                   DLOG(WWDW/T413)
     3+DWWWD*                   DLOG(WWWD/T415)
     4+DWWWW*                   DLOG(WWWW/T415)
     5+WDDDD*                   DLDG(DDDD/T41)
     6+WDDDW*                   DLOG(DDDW/T41)
```

```
     7+WDDWD*                    OLDG(DDWD/T43)
     8+WDDWW*                    DLDG(DDWW/T43)
C
      X2=WDWDD*DLDG(WDWDD/T521)
     &+WDWDW*DLOG(WDWDW/T521)
     1+WDWWD*DLOG(WDWWD/T523)
     2+WDWWW*DLOG(WDWWW/T523)
     3+WWDDD*DLOG(WWDDD/T525)
     4+WWDDW*DLDG(WWDDW/T525)
     4+WWDWD*DLDG(WWDWD/T527)
     5+WWDWW*DLDG(WWDWW/T527)
     5+WWWDD*DLOG(WWWDD/T529)
     6+WWWDW*DLDG(WWWDW/T529)
     7+WWWWD*DLOG(WWWWD/T531)
     7+WWWWW*DLDG(WWWWW/T531)
C
      X4=WDWDD*                  DLDG(DWDD/T45)
     &+WDWDW*                    DLDG(DWDW/T45)
     1+WDWWD*·                   DLOG(DWWD/T47)
     2+WDWWW*                    DLOG(DWWW/T47)
     3+WWDDO*                    OLDG(WDDD/T49)
     4+WWDDW*                    OLOG(WDDW/T49)
     4+WWDWD*                    OLOG(WDWD/T411)
     5+WWDWW*                    DLOG(WDWW/T411)
     5+WWWDD*                    DLOG(WWDD/T413)
     6+WWWDW*                    DLOG(WWDW/T413)
     7+WWWWD*                    DLOG(WWWD/T415)
     7+WWWWW*                    OLOG(WWWW/T415)
C
C...CALCULATE THE STATISTIC
C
      ETA3=2.DO*(X1+X2-(X3+X4))
C
C...STATISTIC TO TEST NULL HYPOTHESIS THAT THE CHAIN
C   IS DF ORDER 2 < 3
C
      Y1=WWWW*DLOG(WWWW/T415)
     1+WWWD* DLOG(WWWD/T415)
     1+DWWW* DLDG(DWWW/T47)
     1+DWWD* DLDG(DWWD/T47)
     3+WDWW* DLOG(WDWW/T411)
     4+WDWO* DLDG(WDWD/T411)
     5+DDWW* DLOG(DDWW/T43)
     6+DDWD* DLDG(DDWD/T43)
     7+WWDW* DLOG(WWDW/T413)
     8+WWDD* DLOG(WWDD/T413)
     9+DWDW* DLOG(DWDW/T45)
     *+DWDD* DLOG(DWDD/T45)
     1+WDOW* DLDG(WDDW/T49)
     2+WDDO* DLDG(WDDD/T49)
     3+ODDW* DLDG(DDDW/T41)
     4+ODOO* DLOG(DDDD/T41)
C
      Y2=WWWW*                   DLOG(WWW/T37)
     1+WWWD*                     DLOG(WWD/T37)
     1+DWWW*                     DLOG(WWW/T37)
     1+DWWD*                     DLDG(WWD/T37)
     3+WDWW*                     DLOG(DWW/T33)
     4+WDWD*                     DLOG(OWD/T33)
     5+DDWW*                     DLDG(DWW/T33)
     6+DDWD*                     DLOG(OWD/T33)
     7+WWDW*                     DLOG(WDW/T35)
     8+WWDD*                     DLOG(WDD/T35)
     9+DWDW*                     DLDG(WDW/T35)
     *+DWDO*                     DLDG(WDD/T35)
     1+WDDW*                     DLDG(DDW/T31)
```

```
     2+WDDD*                        DLDG(DDD/T31)
     3+DDDW*                        DLDG(DDW/T31)
     4+DDDD*                        DLDG(DDD/T31)
C
C...CALCULATE THE STATISTIC
C
      ETA2=2.DO*(Y1-Y2)
C
C...STATISTIC TO TEST NULL HYPDTHESIS THAT CHAIN IS DF
C   DRDER 1 < 2
C
      Z1=WWW*DLOG(WWW/T37)
     1+WWD* DLOG(WWD/T37)
     2+DWW* DLOG(DWW/T33)
     3+DWD* DLDG(DWD/T33)
     4+WDW* DLOG(WDW/T35)
     5+WDD* DLOG(WDD/T35)
     6+DDW* DLOG(DDW/T31)
     7+DDD* DLDG(DDD/T31)
C
      Z2=WWW*                       DLOG(WW/T23)
     1+WWD*                         DLDG(WD/T23)
     2+DWW*                         DLDG(WW/T23)
     3+DWD*                         DLOG(WD/T23)
     4+WDW*                         DLDG(DW/T21)
     5+WDD*                         DLOG(DD/T21)
     6+DDW*                         DLDG(DW/T21)
     7+DDD*                         DLDG(DD/T21)
C
C...CALCULATE THE STATISTIC
C
      ETA1=2.DO*(Z1-Z2)
C
C...STATISTIC TD TEST NULL HYPDTHESIS THAT CHAIN IS DF
C   ORDER O < 1
C
      Z3=WW*DLOG(WW/T23)
     1+WD* DLOG(WD/T23)
     2+DW* DLDG(DW/T21)
     3+DD* DLOG(DD/T21)
C
      Z4=WW*                        DLDG(W/T11)
     1+WD*                          DLDG(D/T11)
     2+DW*                          DLDG(W/T11)
     3+DD*                          DLDG(D/T11)
C
C...CALCULATE THE STATISTIC
C
      ETAO=2.DO*(Z3-Z4)
C
C...TO TEST HYPOTHESIS THAT CHAIN IS OF DRDER
C   K < 4 CALCULATE THE AIC CRITERIDN AS SUGGESTED
C   BY TONG (REF. TONG, 1975; GATES AND TDNG, 1976)
C
      R(1)=ETAO+ETA1+ETA2+ETA3-30.ODO
      R(2)=ETA1+ETA2+ETA3-28.ODO
      R(3)=ETA2+ETA3-24.ODO
      R(4)=ETA3-16.ODO
      R(5)=O.ODO
C
C...CALCULATE SCHWARZ BAYESIAN CRITERION
C   REF(KATZ, 1979A; SCHWARZ 1978)
C
      SB(1)=ETAO+ETA1+ETA2+ETA3-15.ODO*DLDG(T11)
      SB(2)=ETA1+ETA2+ETA3-14.ODO*DLOG(T11)
      SB(3)=ETA2+ETA3-12.ODO*DLOG(T11)
```

```
      SB(4)=ETA3-8.ODO*DLOG(T11)
      SB(5)=O.DO
C
C...OUTPUT THE TOTALS
C
      WRITE(6,240) JB,JE
      WRITE(6,200) W,D,T11
      WRITE(6,200) DW,DD,T21
      WRITE(6,200) WW,WD,T23
      WRITE(6,200) DDW,DDD,T31
      WRITE(6,200) DWW,DWD,T33
      WRITE(6,200) WDW,WDD,T35
      WRITE(6,200) WWW,WWD,T37
      WRITE(6,200) DDDW,DDDD,T41
      WRITE(6,200) DDWW,DDWD,T43
      WRITE(6,200) DWDW,DWDD,T45
      WRITE(6,200) DWWW,DWWD,T47
      WRITE(6,200) WDDW,WDDD,T49
      WRITE(6,200) WDWW,WDWD,T411
      WRITE(6,200) WWDW,WWDD,T413
      WRITE(6,200) WWWW,WWWD,T415
      WRITE(6,200) DDDDW,DDDDD,T51
      WRITE(6,200) DDDWW,DDDWD,T53
      WRITE(6,200) DDWDW,DDWDD,T55
      WRITE(6,200) DDWWW,DDWWD,T57
      WRITE(6,200) DWDDW,DWDDD,T59
      WRITE(6,200) DWDWW,DWDWD,T511
      WRITE(6,200) DWWDW,DWWDD,T513
      WRITE(6,200) DWWWW,DWWWD,T515
      WRITE(6,200) WDDDW,WDDDD,T517
      WRITE(6,200) WDDWW,WDDWD,T519
      WRITE(6,200) WDWDW,WDWDD,T521
      WRITE(6,200) WDWWW,WDWWD,T523
      WRITE(6,200) WWDDW,WWDDD,T525
      WRITE(6,200) WWDWW,WWDWD,T527
      WRITE(6,200) WWWDW,WWWDD,T529
      WRITE(6,200) WWWWW,WWWWD,T531
      WRITE(6,230) ETAO,ETA1,ETA2,ETA3
      WRITE(6,220)
C
C...OUTPUT THE AIC, SBC CRITERION
C
      DO 20 I=1,5
      J=I-1
 20   WRITE(6,210) J,R(I),SB(I)
 50   CONTINUE
      RETURN
C
C...FORMAT STATEMENTS
C
 200  FORMAT(1X,3F8.0)
 210  FORMAT(1X,I6,10X,F9.3,F11.3)
 220  FORMAT('O',5X,'K',13X,'AIC(K)',5X,'SBC(K)')
 230  FORMAT('O','THE MAXIMUM LIKELIHOOD RATIO TEST',
     &' STATISTICS ARE O TO 3'/1X,4F11.4)
 240  FORMAT('OSEQUENCE TOTALS FOR TIME PERIOD',2I5)
      END
```

```
      SUBROUTINE FOUR(COUNT1,COUNT2,COUNT3,COUNT4,COUNT5)
C
C...SUBROUTINE FOUR,  CREATED 79 10 11
C   LAST MODIFIED 80 05 19
C
C...SUBROUTINE CALCULATES FIRST 21 HARMONICS OF THE FOURIER
C   SERIES APPROXIMATION TO THE PROBABILITY OF DRY DAYS
C   CALCULATES THE CUMULATIVE PERIODOGRAM ESTIMATES AND PLOT
C   THE PERIODOGRAMS  (REF. YEVYEVICH,1972)
C
C...ROUTINES REQUIRED   *PLOTLIB(SYSTEM SUBROUTINE U OF A)
C
C...I/O   6=CUMULATIVE PERIODOGRAM VALUES
C         9=PLOT FILE
C        10=OUTPUT, FOURIER COEFFICIENTS
C
C...INITIALIZATION
C
      IMPLICIT REAL*8 (A-B,D-H,O-Z)
      LOGICAL*1 LFMT(1)/'*'/
      REAL*4 PVAR(30),HAR(30),LEGEND(31),AB(2),ORD(4)
      INTEGER COUNT1(365,2),COUNT2(365,4),COUNT3(365,8),
     1        COUNT4(365,16), COUNT5(365,32),PFLAG
      DIMENSION A(21,31),B(21,31),AMP(21,31),PHI(21,31),
     1          ASPBS(21,31),   VAR(31)
      DATA PI/3.14159 26535 89793/,ONE/1.DO/,
     5     AB/'HARM','ONIC'/,ORD/' EXP',
     6     'LAIN','ED  ','VAR '/
     7     ,VAR/31*O.DO/,PFLAG/1/
C PLOTTING INITIALIZATION
      CALL PLOTS
      CALL FACTOR(0.6)
      CALL ORGEP(1.5,1.5,2.0)
C
C...BEGIN TABULATIONS FOR FOURIER COEFFICIENTS
C
      DO 200 J=1,21
          M=J-1
          PREFIX=2.DO/365.DO
          DO 10 K=1,31
              A(J,K)=O.DO
   10       B(J,K)=O.DO
          DO 180 I=1,365
              T11=DFLOAT(COUNT1(I,1)+COUNT1(I,2))
              T21=DFLOAT(COUNT2(I,1)+COUNT2(I,2))
              T23=DFLOAT(COUNT2(I,3)+COUNT2(I,4))
              T31=DFLOAT(COUNT3(I,1)+COUNT3(I,2))
              T33=DFLOAT(COUNT3(I,3)+COUNT3(I,4))
              T35=DFLOAT(COUNT3(I,5)+COUNT3(I,6))
              T37=DFLOAT(COUNT3(I,7)+COUNT3(I,8))
C
C...BEGIN TABULATIONS, CHECKING FOR ZERO TOTALS
C
              ANG=PI*PREFIX*DFLOAT(M*I)
              DC=DCOS(ANG)
              DS=DSIN(ANG)
              IF(T11.LT.ONE) GO TO 30
              A(J,1)=A(J,1)+DFLOAT(COUNT1(I,1))/T11*DC
              B(J,1)=B(J,1)+DFLOAT(COUNT1(I,1))/T11*DS
              IF(M.EQ.O) VAR(1)=VAR(1)+(DFLOAT(COUNT1(I,1))/
     1        T11)**2
   30         IF(T21.LT.ONE) GO TO 40
              A(J,2)=A(J,2)+DFLOAT(COUNT2(I,1))/T21*DC
```

```
             B(J,2)=B(J,2)+DFLOAT(COUNT2(I,1))/T21*DS
             IF(M.EQ.O) VAR(2)=VAR(2)+(DFLOAT(COUNT2(I,1))/
     1      T21)**2
   40        IF(T23.LT.ONE) GO TO 50
             A(J,3)=A(J,3)+DFLOAT(COUNT2(I,3))/T23*DC
             B(J,3)=B(J,3)+DFLOAT(COUNT2(I,3))/T23*DS
             IF(M.EQ.O) VAR(3)=VAR(3)+(DFLOAT(COUNT2(I,3))/
     1      T23)**2
   50        IF(T31.LT.ONE) GO TO 60
             A(J,4)=A(J,4)+DFLOAT(COUNT3(I,1))/T31*DC
             B(J,4)=B(J,4)+DFLOAT(COUNT3(I,1))/T31*DS
             IF(M.EQ.O) VAR(4)=VAR(4)+(DFLOAT(COUNT3(I,1))/
     1      T31)**2
   60        IF(T33.LT.ONE) GO TO 70
             A(J,5)=A(J,5)+DFLOAT(COUNT3(I,3))/T33*DC
             B(J,5)=B(J,5)+DFLOAT(COUNT3(I,3))/T33*DS
             IF(M.EQ.O) VAR(5)=VAR(5)+(DFLOAT(COUNT3(I,3))/
     1      T33)**2
   70        IF(T35.LT.ONE) GO TO 80
             A(J,6)=A(J,6)+DFLOAT(COUNT3(I,5))/T35*DC
             B(J,6)=B(J,6)+DFLOAT(COUNT3(I,5))/T35*DS
             IF(M.EQ.O) VAR(6)=VAR(6)+(DFLOAT(COUNT3(I,5))/
     1      T35)**2
   80        IF(T37.LT.ONE) GO TO 170
             A(J,7)=A(J,7)+DFLOAT(COUNT3(I,7))/T37*DC
             B(J,7)=B(J,7)+DFLOAT(COUNT3(I,7))/T37*DS
             IF(M.EQ.O) VAR(7)=VAR(7)+(DFLOAT(COUNT3(I,7))/
     1      T37)**2
  170       CONTINUE
C
C
C CALCULATE COEFFICIENTS
C
  180       CONTINUE
            IF(M.EQ.O) PREFIX=PREFIX/2.DO
            DO 190 K=1,7
             A(J,K)=PREFIX*A(J,K)
             B(J,K)=PREFIX*B(J,K)
             ASPBS(J,K)=A(J,K)*A(J,K)+B(J,K)*B(J,K)
             AMP(J,K)=DSQRT(ASPBS(J,K))
             PHI(J,K)=DATAN(-1.DO*B(J,K)/A(J,K))
  190         ASPBS(J,K)=ASPBS(J,K)/2.DO
  200     CONTINUE
        WRITE(6,260)
C...CALCULATE AND OUTPUT CUMULATIVE PERIODOGRAM
        DO 230 K=1,7
              VAR(K)=VAR(K)/365.DO-A(1,K)*A(1,K)
               VAR(K)=VAR(K)*365.DO/364.DO
            SUM=O.DO
            WRITE(6,270) K
            DO 210 I=2,21
             M=I-1
             SUM=SUM+ASPBS(I,K)
             HAR(M)=FLOAT(M)
             PVAR(M)=SUM/VAR(K)
  210         WRITE(6,280) M,PVAR(M)
C...PLOT CUMULATIVE PERIODOGRAM
          IF(PFLAG.NE.1) GO TO 220
            CALL KJPL(HAR,PVAR,20,2.5,AB,ORD,1.6,PFLAG)
          IF(K.NE.1) PFLAG=2
          GO TO 230
  220       CALL KJPL(HAR,PVAR,20,2.5,AB,ORD,1.6,PFLAG)
          PFLAG=1
  230 CONTINUE
      DO 250 K=1,7
C...OUTPUT COEFFICIENTS
```

```
          WRITE(10,LFMT) VAR(K)
            DO 240 I=1,21
              M=I-1
  240       WRITE(10,290) M,A(I,K),B(I,K),AMP(I,K),PHI(I,K)
  250 CONTINUE
      CALL PLOT(0.,0.,999)
      RETURN
  260 FORMAT('1',10X,'CUMULATIVE PERIODOGRAM VALUES')
  270   FORMAT('0',10X,'SEQUENCE NUMBER',I3/21X,'M',10X,
     1         'PVAR')
  280       FORMAT(20X,I2,9X,F6.4)
  290 FORMAT(1X,I3,4F10.6)
      END
C
C
      SUBROUTINE KJPL(X,Y,NM,XSC,
     &ABS,ORD,DTIC,PFLAG)
C...THIS ROUTINE PLOTS THE CUMULATIVE PERIODOGRAM USING
C   U OF A SYSTEM PLOTTING ROUTINES
C
      DIMENSION X(30),Y(30),ABS(2),ORD(4)
      INTEGER PFLAG
      X(NM+1)=0.
      X(NM+2)=XSC
      Y(NM+1)=0.
      Y(NM+2)=0.2
C
C...IF CALCOMPQ TO BE USED THE ORIGIN CAN BE CHANGED
C   BY CHANGING THE NUMBER OF PLOTS IN POSITION 1 OF THE
C   CALL TO ORIGIN TO 4
C
      CALL LINEP(0.15)
      IF(PFLAG.NE.1) GO TO 10
      CALL  ORIGIN(2,9.,6.,.75,.75)
      CALL AX2EP(1.,3,1,1,1.3)
      CALL AXIS2(0.,0.,ORD,16,5.1,90.,0.,.2,-1.)
      CALL AXIS2(8.3,0.,' ',1,-5.1,90.,0.,.2,1.)
      CALL AX2EP(1.,3,0,0,1.3)
      CALL AXIS2(0.,0.,ABS,-8,8.3,0.,0.,XSC,DTIC)
      CALL AXIS2(0.,5.1,' ',-1,-8.3,0.,0.,XSC,DTIC)
      CALL LINE(X,Y,NM,1,-1,2)
      GO TO 20
   10 CALL LINE(X,Y,NM,1,-1,3)
   20 RETURN
      END
```

```
C           TW MARKOV CHAIN EXPONENTIAL MODEL 09/10/71
C
C VERSION FROM TODOROVIC AND WOOLHISER (1974)
C
C I/O 5=INPUT, 6=OUTPUT DIAGNOSTICS 7=RESULTS
C
C THIS PROGRAM COMPUTES THE CDF FOR TOTAL RAINFALL FOR N DAYS.  INPUT
C PARAMETERS ARE N, Q0=P01=P(DAY I IS WET GIVEN DAY I-1 IS DRY), QI=P11=P(DAY
C I IS WET GIVEN DAY I-1 IS WET), XLAM IS PARAMETER IN NEG. EXPONENTIAL
C DISTRIBUTION, R=P=P(DAY BEFORE SERIES BEGINS IS WET)
      LOGICAL*1 LFMT(1) /'*'/
      DIMENSION PSIO(50), PSII(50), PSI(50), G(300), H(300), XG(300),
     1          XH(300)
   10 READ (5,LFMT,END=160) N, Q0, QI, R, XLAM, XGD, NG, XHD, NH
   20 WRITE (6,170) N, Q0, QI, R, XLAM
C
C  THIS SECTION COMPUTES THE CONDITIONAL COUNTING PROCESS DENSITY
C FUNCTIONS (REF. GABRIEL, 1959)
C PSIO(I),PSII(I), WHERE I=NU+1
      PSII(1) = (1. - Q0) ** (N - 1) * (1. - QI)
      PSIO(1) = (1. - Q0) ** N
      NU = 0
      NT = N + 1
      DO 90 I = 2, N
        NU = NU + 1
        NCI = IFIX(N + 0.5 - ABS(2*NU - N + 0.5) + 0.01)
        NCO = IFIX(N + 0.5 - ABS(2*NU - N - 0.5) + 0.01)
        NC = 0
        A = 0.0
        B = 1.
        NSW = 1
        SUMI = 0.0
        SUMO = 0.0
        TERMI = (1. - QI) / (1. - Q0)
        TERMO = Q0 / QI
   30   SUMI = SUMI + TERMI
        SUMO = SUMO + TERMO
        NC = NC + 1
        IF (NC .NE. NCO) GO TO 60
        PSIO(I) = QI ** NU * (1. - Q0) ** (N - NU) * SUMO
        IF (NCO .GT. NCI) GO TO 90
        IF (NC .EQ. NCI) GO TO 70
   40   GO TO (50, 80), NSW
   50   NSW = 2
        A = A + 1.0
        TERMI = TERMI * (NU - A + 1.) / A * Q0 / QI
        TERMO = TERMO * (N - NU - A + 1.) / A * (1. - QI) / (1. - Q0)
        GO TO 30
   60   IF (NC .NE. NCI) GO TO 40
   70   PSII(I) = QI ** NU * (1. - Q0) ** (N - NU) * SUMI
        IF (NCO .GT. NCI) GO TO 40
        GO TO 90
   80   NSW = 1
        B = B + 1
        TERMI = TERMI * (N - NU - B + 1.) / (B - 1.) * (1. - QI) / (1. -
     1  Q0)
        TERMO = TERMO * (NU - B + 1.) / (B - 1.) * Q0 / QI
        GO TO 30
   90 CONTINUE
  100 PSII(NT) = QI ** N
      PSIO(NT) = QI ** (N - 1) * Q0
      WRITE (7,LFMT) NT
      S=0.0
```

```
      DO 110 I = 1, NT
      DAY=FLOAT(I-1)
        PSI(I) = R * PSII(I) + (1. - R) * PSIO(I)
      S=S+PSI(I)
  110 WRITE (7,180) DAY, S, PSI(I), PSIO(I), PSII(I)
C
C...GABRIEL'S METHOD OF CALCULATING THE DISTN IS COMPLETE
C
C THIS SECTION COMPUTES THE CDF OF THE MAXIMUM DAILY RAINFALL FOR N DAYS
      WRITE (7,LFMT) NG
      XG(1) = 0.0
      DO 130 I = 1, NG
        G(I) = 0.0
        FP = 1.0 - EXP(-1.0*XLAM*XG(I))
        DO 120 J = 2, NT
C
C   NOTE K FROM 1 TO N, J FROM 2 TO NT, K'TH WET DAY IN PSI(K+1)
C
          K = J - 1
  120   G(I) = G(I) + FP ** FLOAT(K) * PSI(J)
        G(I) = G(I) + PSI(1)
        XG(I + 1) = XG(I) + XGD
  130 WRITE (7,180) XG(I), G(I)
C
C... THIS SECTION COMPUTES THE CDF FOR THE TOTAL PCPN IN N DAYS
C
      WRITE (7,LFMT) NH
      XH(1) = 0.0
      DO 150 I = 1, NH
        H(I) = 0.0
        DO 140 J = 2, NT
          K = J - 1
          RK = FLOAT(K)
C
C...EVALUATE P(XK<X) USING IMSL (1979) ROUTINES
C
          CALL GAM(XH(I), F, RK, XLAM)
  140   H(I) = H(I) + F * PSI(J)
        H(I) = H(I) + PSI(1)
        XH(I + 1) = XH(I) + XHD
  150 WRITE (7,180) XH(I), H(I)
      GO TO 10
  160 STOP
  170 FORMAT ('0', 10X, 'N=', I3, ' QO=', F5.4, ' QI=', F5.4, ' R=',
     1       F5.4, ' LAMDA=', F5.2)
  180 FORMAT (1X, F6.1,4F7.3)
      END
```

```
C        KATZ DISTRIBUTIONS
C
C...CREATED BY K. JOHNSTONE 79/11/14
C
C...LAST MODIFIED 80/01/23
C
C...PROGRAM USES METHOD OF KATZ (REF. 1974, 1977)
C   TO GENERATE THE DISTRIBUTION OF THE NUMBER OF
C   WET DAYS IN N DAYS AND THE MAXIMUM DAILY PCPN
C   IN N DAYS, AND THE TOTAL PRECIPITATION IN N DAYS.
C
C...PARAMETERS
C       WO,W1,W  ARRAYS CONTAINING THE DISTRIBUTION OF THE NUMBER
C         OF WET DAYS IN A TOTAL OF N
C       P  PROBABILITY OF A WET DAY
C       OMP  PROBABILITY OF A DRY DAY
C       PO1,P11,P10,POO  TRANSITION PROBABILITIES WHERE O REPRESENTS
C         A WET AND 1 A DRY DAY; OBTAINED FROM FSG
C       N  NUMBER OF DAYS THE DISTRIBUTION IS REQUIRED FOR, MAX IS 31
C       INITD  INITIAL DAY IN YEAR FOR WHICH DISTRIBUTION IS CALCULATED
C...INITITALIZATION AND DIMENSIONING
C
C...SUBROUTINES...FSG=GENERATES TRANSITION PROBS GIVEN
C                      FOURIER SERIES COEFFICIENTS
C
C                 MAXP=GENERATES DISTRIBUTION FOR MAXIMUM DAILY
C                      PRECIPITATION IN N DAYS
C                 TOTP=GENERATES DISTRIBUTION OF TOTAL
C                      PRECIPITATION IN N DAYS
C                 GAM=INTEGRATES THE GAMMA DENSITY FUNCTION
C                 DERIV=DIRRENTIATES THE DISTRIBUTIONS TO OBTAIN
C                       DENSITY FUNCTIONS
C                 SIMPS=USES SIMPSONS RULE TO DO THE CONVOLUTION
C                        INTEGRATIONS
C
      IMPLICIT REAL*8 (A-H,O-Z)
      LOGICAL*1 LFMT(1)/'*'/
      DIMENSION WO(35,35),W1(35,35),P(32),OMP(32),PO1(32),
     &          P11(32),P10(32),POO(32)
      COMMON P,OMP,PO1,P11,P10,POO
      DO 20 I=1,35
      DO 10 J=1,35
      WO(J,I)=0.DO
 10   W1(J,I)=0.DO
 20   CONTINUE
      WO(2,2)=1.DO
      W1(2,2)=1.DO
      WRITE(6,30)
 30   FORMAT('OENTER THE NUMBER OF DAYS, THE BEGINNING DAY',
     &' FOR THE DISTRIBUTION')
C
C...IT IS REQUIRED THAT N BE 1 LESS THAN THE TOTAL TIME
C   PERIOD THAT IS CONSIDERED
C
      READ(5,LFMT) N,INITD
C
C...GENERATE THE TRANSITION AND INITIAL PROBABILITIES
C
      CALL FSG(INITD,N)
C
C...CALCULATE THE DISTRIBUTIONS FOR WET AND DRY DAY OCCURRENCES
C
      L=N+2
```

```
      DO 50 J=3,L
      DO 40 I=2,J
      K=N-J+4
      WO(I,J)=POO(K)*WO(I,J-1)+PO1(K)*W1(I-1,J-1)
 40   W1(I,J)=P10(K)*WO(I,J-1)+P11(K)*W1(I-1,J-1)
 50   CONTINUE
C
C...OUTPUT HEADER, THEN CALCULATE AND OUTPUT FINAL DIST.
C
      WRITE(7,70) N
 70   FORMAT('1TOTAL NUMBER OF DAYS',I3/'O',4X,'K',5X,'WD',5X,'W',
     &6X,'WO',5X,'W1')
      WTOT=O.
      DO 60 I=2,L
      W=OMP(1)*WO(I,L)+P(1)*W1(I,L)
      WTOT=WTOT+W
      DAYW=FLOAT(I-2)
 60   WRITE(7,80) DAYW,WTOT,W,WO(I,L),W1(I,L)
 80   FORMAT(1X,F6.1,4F7.3)
      CALL MAXP(N)
      STOP
      END
```

```
-       SUBROUTINE MAXP(N)
C
C...THIS ROUTINE CALCULATES THE DISTRIBUTION OF THE MAXIMUM
C   PRECIPITATION IN N DAYS
C
C...PARAMETERS
C         X  ARRAY OF X VALUES
C         Y ARRAY OF DISTRIBUTION VALUES
C         DY  ARRAY OF DENSITY FUNCTION VALUES
C         P'S INITIAL AND TRANSITION PROBABILITIES
C         LAMDAO  SCALE PARAMETER FOR GAMMA DISTRIBUTION WHEN
C           PREVIOUS DAY IS DRY
C         LAMDA1  SCALE PARAMETER FOR GAMMA DISTRIBUTION WHEN
C           PREVIOUS DAY IS WET
C         ETA'S  SHAPE PARAMETERS FOR THE GAMMA DISTRIBUTION
C         NX  NUMBER OF X VALUES
C         DELX  X(I)-X(I-1)
C         FO,F1  PROBABILITY THAT AMOUNT OF PRECIPITATION <=X
C           FOR PREVIOUS DAY DRY OR WET RESPECTIVELY
C         GO,G1,GOT  DISTRIBUTION VALUE FOR A PARTICULAR X
C
C...INITIALIZATION AND DIMENSIONING
C
        IMPLICIT REAL*8 (A-H,O-Z)
        DIMENSION X(200),DY(200),Y(200)
        DIMENSION OMP(32),P(32),PO1(32),POO(32),P10(32),P11(32)
        COMMON P,OMP,PO1,P11,P10,POO
        LOGICAL*1 LFMT(1)/'*'/
        REAL*8 LAMDAO,LAMDA1
C
C...INPUT DELTA X AND GAMMA DISTRIBUTION PARAMETERS
C
        WRITE(6,10)
 10     FORMAT('OENTER NUMBER OF X''S AND DELTA X')
        READ(5,LFMT) NX,DELX
        X(1)=O.DO
        DO 12 IM=2,NX
 12     X(IM)=X(IM-1)+DELX
        WRITE(6,20)
 20     FORMAT('OENTER  ETAO, LAMDAO')
        READ(5,LFMT) ETAO,LAMDAO
        WRITE(6,25)
 25     FORMAT('OENTER ETA1,LAMDA1')
        READ(5,LFMT) ETA1,LAMDA1
C
C...WRITE HEADER FOR OUTPUT
        WRITE(7,60)
 60     FORMAT('O',4X,'X',5X,'G',6X,'DG')
C
C...CALCULATE DISTRIBUTION VALUE FOR EACH X DESIRED
C
        DO 50 IM=1,NX
C
C...DETERMINE F(X)=PROB(XOBS<=X) FOR GAMMA VARIATE
C
        CALL GAM(X(IM),FO,ETAO,LAMDAO)
        CALL GAM(X(IM),F1,ETA1,LAMDA1)
        GO=1.DO
        G1=1.DO
C
C...BEGIN CALCULATION FOR VALUE X
C
        DO 30 KM=1,N
```

```
C
C...CALCULATE INDEX OF THE TRANSITION PROBABILITIES
C   SO THAT THEY PROGRESS FROM N+1 TO 2, THE N DAYS
C   THE DISTRIBUTION IS CALCULATED FOR, INDEX=1
C   CORRESPONDS TO DAY=0
C
      JM=N-KM+2
      GOT=P00(JM)*G0+P01(JM)*F0*G1
      G1=P10(JM)*G0+P11(JM)*F1*G1
 30   G0=GOT
C
C...DETERMINE THE FINAL DISTRIBUTION VALUE FOR AN X
C
      G=OMP(1)*G0+P(1)*G1
C
C...SAVE DISTRIBUTION VALUES IN ORDER TO CALCULATE THE
C   DENSITY FUNCTION
C
 50   Y(IM)=G
C
C...CALCULATE THE DENSITY FUNCTION
C
      CALL DERIV(Y,DY,NX,DELX,200)
C
C...OUTPUT RESULTS
C
      DO 500 IM=1,NX
 500  WRITE(7,40) X(IM),Y(IM),DY(IM)
 40   FORMAT(1X,F6.1,2F7.3)
      CALL TOTP(N,ETA0,LAMDA0,ETA1,LAMDA1)
      RETURN
      END
```

```
      SUBROUTINE TOTP(N,ETAO,LAMOAO,ETA1,LAMDA1)
C
C...ROUTINE CREATEO NOVEMBER 1979
C...LAST MOOIFIED 80 01 21
C
C...ROUTINE CALCULATES THE OISTRIBUTION FUNCTION FOR THE
C   TOTAL AMOUNT OF PRECIPITATION IN N OAYS
C
C...PARAMETERS
C         P'S  INITIAL ANO TRANSITION PROBABILITIES
C         TX  ARRAY OF X VALUES
C         HO,H1  CONDITIONAL OISTRIBUTION FUNCTIONS
C         HDO,HD1  DUMMY ARRAYS
C         CONO,CON1  ARRAYS OF CONVOLUTEO PROOUCTS
C         CO,C1  CONVOLUTION RESULTS
C         FPO,FP1  GAMMA DENSITY FUNCTIONS
C         NXT  NUMBER OF X VALUES
C         OELXT  OELTA X
C         NUMO  OUTPUT INCREMENT
C         E'S  SHAPE PARAMETERS FOR INTEGRALS FROM O TO DELXT
C           AND 2 OELXT
C         PROB'S 'CONVOLUTION' FROM ZERO TO OELXT
C         PR'S  'CONVOLUTION' FROM ZERO TO 2 OELXT
C         A1,B1 COEFFICIENTS FOR FIRST ORDER POLYNOMIAL
C         A2,B2,C2  COEFFICIENTS FOR SECOND ORDER POLYNOMIAL
C         A,B,C  COEFFICIENTS FOR SECOND OROER POLYNOMIAL
C
C...DIMENSIONING
C
      IMPLICIT REAL*8 (A-H,O-Z)
      LOGICAL*1 LFMT(1)/'*'/
      REAL*8 LAMOAO,LAMDA1
      DIMENSION P(32),OMP(32),PO1(32),P11(32),P10(32),POO(32),
     1TX(1000),HO(1000),H1(1000),
     2HDO(1000),HO1(1000),CONO(1000),CON1(1000),
     3FPO(1000),FP1(1000)
      COMMON P,OMP,PO1,P11,P10,POO
C
C...INPUT VALUES FOR RANGE, OELTA X AND OUTPUT INTERVAL
C
      WRITE(6,10)
 10   FORMAT('OENTER NUMBER OF X''S, DELTA X AND INCREMENT FOR OUTPUT')
      REAO(5,LFMT) NXT,DELXT,NUMO
      IF(NXT.EQ.O) RETURN
C
C...INITIALIZATION
C
      TX(1)=O.DO
      DO 20 IT=2,NXT
 20   TX(IT)=TX(IT-1)+DELXT
      DSQ=DELXT*DELXT
      EO2=ETAO+2.OO
      EO1=ETAO+1.OO
      E12=ETA1+2.DO
      E11=ETA1+1.DO
C
C...TO EVALUATE CONVOLUTIONS FROM O TO X IT IS IMPOSSIBLE
C   TO USE SIMPSONS RULE ALONE BECAUSE THE GAMMA OENSITY
C   FUNCTION TENDS TO INFINITY FOR X GOING TO O.
C   FIT A 1ST ANO 2ND OROER LAGRANGIAN POLYNOMIAL
C   TO REFLECTION OF FIRST FEW GRIO POINTS OF DISTN.
C   FOR TOTAL PRECIP.  THEN USE GAM TO EVALUATE CONVOL.
C   INTEGRAL OVER RANGE O TO DELXT OR 2DELXT, CORRECT
```

```
C     FOR NEW ORDER OF GAMMA DENSITY FN., MULTIPLY BY
C     POLYNOMIAL COEFF. TO OBTAIN CONVOL. FROM O TO OELXT
C     OR 2OELXT. AOD VALUE TO THE CONVOLUTION OBTAINEO USING
C     SIMPSONS RULE OVER THE REMAINING RANGE OF INTEGRATION.
C
C...OBTAIN VALUES FOR THE INCOMPLETE GAMMA FUNCTION
C     AT DELXT AND 2 DELXT
C
      CALL GAM(TX(2),PROBO2,EO2,LAMDAO)
      CALL GAM(TX(2),PROBO1,EO1,LAMOAO)
      CALL GAM(TX(2),PROB12,E12,LAMDA1)
      CALL GAM(TX(2),PROB11,E11,LAMDA1)
      CALL GAM(TX(2),PROB1,ETA1,LAMOA1)
      CALL GAM(TX(2),PROBO,ETAO,LAMDAO)
      CALL GAM(TX(3),PRO2,EO2,LAMOAO)
      CALL GAM(TX(3),PRO1,EO1,LAMOAO)
      CALL GAM(TX(3),PR12,E12,LAMOA1)
      CALL GAM(TX(3),PR11,E11,LAMDA1)
      CALL GAM(TX(3),PR1,ETA1,LAMOA1)
      CALL GAM(TX(3),PRO,ETAO,LAMOAO)
C
C...CORRECT VALUES OF THE INCOMPLETE GAMMA FUNCTION TO
C     OBTAIN INTEGRAL OF X TO SOME POWER TIMES THE GAMMA
C     DENSITY FUNCTION FROM ZERO TO DELXT ANO 2 DELXT
C
      G1=OGAMMA(ETA1)
      GO=OGAMMA(ETAO)
      GO2=DGAMMA(EO2)
      GO1=DGAMMA(EO1)
      G12=DGAMMA(E12)
      G11=DGAMMA(E11)
      PROBO2=PROBO2*GO2/(GO*LAMDAO*LAMOAO)
      PROBO1=PROBO1*GO1/(GO*LAMOAO)
      PROB12=PROB12*G12/(G1*LAMDA1*LAMDA1)
      PROB11=PROB11*G11/(G1*LAMDA1)
      PRO2=PRO2*GO2/(GO*LAMDAO*LAMDAO)
      PRO1=PRO1*GO1/(GO*LAMDAO)
      PR12=PR12*G12/(G1*LAMDA1*LAMOA1)
      PR11=PR11*G11/(G1*LAMDA1)
C
C...SET THE DISTRIBUTION FUNCTION TO THEIR INITIAL VALUES
C
      DO 21 IT=1,NXT
 21   HO(IT)=1.DO
      DO 22 IT=1,NXT
 22   H1(IT)=1.DO
C
C...CALCULATE THE VALUES OF THE GAMMA OENSITY FUNCTION
C
      DO 26 IT=2,NXT
        FPO(IT)=LAMDAO**ETAO*TX(IT)**(ETAO-1.DO)
     1 *DEXP(-1.OO*LAMOAO*TX(IT))/GO
        FP1(IT)=LAMOA1**ETA1*TX(IT)**(ETA1-1.DO)
     1 *DEXP(-1.DO*LAMDA1*TX(IT))/G1
 26   CONTINUE
C
C...BEGIN THE CALCULATIONS
C
C...WITH H'S INITIALIZEO TO ONE'S NEED TIME STEPS
C     FROM OAY 1 TO DAY N
      DO 60 IT=1,N
C
C...DETERMINE THE INDEX FOR THE DAILY TRANSITION PROBABILITIES
C
      JT=N-IT+2
C
```

```
C...EVALUATE NEW DISTRIBUTION VALUES FOR ZERO PRECIPITATION
C
      HDO(1)=POO(JT)*HO(1)
      HD1(1)=P10(JT)*HO(1)
C
C...FIT LINE TO FIRST TWO VALUES OF THE REFLECTED H1
C   IN ORDER TO INTEGRATE FROM ZERO TO DELXT
C   AND OBTAIN NEW DISTRIBUTION VALUES FOR DELXT
C
      A1=(H1(1)-H1(2))/DELXT
      B1=H1(2)
      CO=A1*PROBO1+B1*PROBO
      C1=A1*PROB11+B1*PROB1
      HDO(2)=POO(JT)*HO(2)+PO1(JT)*CO
      HD1(2)=P10(JT)*HO(2)+P11(JT)*C1
C
C...FIT QUADRATIC TO REFLECTED H1 IN ORDER TO INTEGRATE FROM ZERO
C   TO 2 DELXT AND OBTAIN THE NEW DISTRIBUTION VALUES
      A2=((H1(3)+H1(1))/2.DO-H1(2))/DSQ
      B2=(H1(2)*TX(3)-(H1(3)*(TX(2)+TX(3))
     & +H1(1)*TX(2))/2.DO)/DSQ
      C2=H1(3)*TX(2)*TX(3)/2.DO/DSQ
      CO=A2*PRO2+B2*PRO1+C2*PRO
      C1=A2*PR12+B2*PR11+C2*PR1
      HDO(3)=POO(JT)*HO(3)+PO1(JT)*CO
      HD1(3)=P10(JT)*HO(3)+P11(JT)*C1
C
C...BEGIN EVALUATING CONVOLUTION FOR REMAINING POINTS
C
      DO 40 KT=4,NXT
      KTM1=KT-1
C
C...EVALUATE CONVOLUTED PRODUCTS PRIOR TO INTEGRATION
      DO 30 KT1=2,KT
      KT1M1=KT1-1
      KT2=KT-KT1+1
      CONO(KT1M1)=FPO(KT1)*H1(KT2)
 30   CON1(KT1M1)=FP1(KT1)*H1(KT2)
C
C...PERFORM INTEGRATIONS USING SIMPSONS RULES
C
      CALL SIMPS(CONO,KTM1,DELXT,CO)
      CALL SIMPS(CON1,KTM1,DELXT,C1)
C
C...DETERMINE COEFFICIENTS FOR LAGRANGIAN POLYNOMIAL
C   WHICH FITS THE REFLECTED H1 VALUES;  IN ORDER TO
C   EVALUATE THE CONVOLUTION NEAR ZERO
C...NOTE THAT TX(1)=O. SO X1 IN THE POLYNOMIAL DERIVATION
C   IS ZERO
C
      A=((H1(KT)+H1(KT-2))/2.DO-H1(KT-1))/DSQ
      B=(H1(KT-1)*TX(3)-(H1(KT)*(TX(2)+TX(3))
     & +H1(KT-2)*TX(2))/2.DO)/DSQ
      C=H1(KT)*TX(2)*TX(3)/2.DO/DSQ
C
C...SUM THE PORTION OF INTEGRATION FROM ZERO TO DELXT
C
      CO=CO+A*PROBO2+B*PROBO1+C*PROBO
      C1=C1+A*PROB12+B*PROB11+C*PROB1
C
C...EVALUATE THE NEW HDO AND HD1 VALUES
C
      HDO(KT)=POO(JT)*HO(KT)+PO1(JT)*CO
 40   HD1(KT)=P10(JT)*HO(KT)+P11(JT)*C1
C
C...UPDATE THE HO AND H1 DISTRIBUTION FUNCTIONS
```

```
C
      DO 50 KT=1,NXT
      HO(KT)=HDO(KT)
 50   H1(KT)=HD1(KT)
 60   CONTINUE
C
C...WHEN THE N DAY DISTRIBUTION FUNCTIONS FOR HO AND H1
C   ARE OBTAINED CALCULATE THE OVERALL DISTRIBUTION FUNCTION
C
      DO 70 KT=1,NXT
 70   HDO(KT)=OMP(1)*HO(KT)+P(1)*H1(KT)
C
C...EVALUATE THE DENSITY FUNCTION FOR THE TOTAL AMOUNT OF PREC.
C
      CALL DERIV(HDO,HD1,NXT,DELXT,1000)
C
C...OUTPUT HEADER AND RESULTS
C
      WRITE(7,80)
 80   FORMAT('O',4X,'X',5X,'H',6X,'DH')
      DO 90 KT=1,NXT,NUMO
 90   WRITE(7,100) TX(KT),HDO(KT),HD1(KT)
 100  FORMAT(1X,F6.1,2F7.3)
      RETURN
      END
```

```
      SUBROUTINE GAM(X,P,ETA,LAMDA)
C
C...THIS ROUTINE CALCULATES THE PROBABILITY F(X)=PROB(XOBS<=X)
C   WHERE IT IS ASSUMED X IS DISTRIBUTED AS A GAMMA VARIATE
C...IT EVALUATES THE INTEGRAL OF (LAMDA**ETA)*(T**(ETA-1))
C   *EXP(-1*LAMDA*T) FROM ZERO TO X
C...THIS IS DONE BY SUBSTITUTING T1=LAMDA *X AND INTEGRATING
C   FROM ZERO TO Y=LAMDA*X USING THE IMSL (1979) ROUTINE MDGAM.
C...VARIABLE ARE CONVERTED TO SINGLE PRECISION PRIOR TO
C   CALLING MDGAM.
C...PARAMETERS
C       X  INPUT UPPER LIMIT TO INTEGRATION
C       P  OUTPUT PROBABILITY F(X)
C       ETA  INPUT SHAPE PARAMETER FOR GAMMA DISTRIBUTION
C       LAMDA  INPUT SCALE PARAMETER FOR GAMMA DISTRIBUTION
C       VARIABLES APPENDED BY AN 'S' ARE IN SINGLE PRECISION
C       IER  OUTPUT ERROR PARAMETER FROM MDGAM, SEE IMSL
C         SUBROUTINE DOCUMENTATION
C
      REAL*8 LAMDA,ETA,X,P,Y
      Y=LAMDA*X
      YS=SNGL(Y)
      ETAS=SNGL(ETA)
      CALL MDGAM(YS,ETAS,PS,IER)
      P=DBLE(PS)
      IF(IER.GT.128) WRITE(6,10) IER,X
 10   FORMAT('OIER= ',I4,' X= ',F5.1)
      RETURN
      END
```

```
      SUBROUTINE FSG(INITD,NUM)
C...PROGRAM FSG (FOURIER SERIES GENERATOR) CREATED  79 11 06
C...LAST MODIFIED 80 02 26
C
C...ROUTINE GENERATES THE FOURIER TIME SERIES APPROXIMATION
C   TO THE PROBABILITY OF THE SELECTED DRY AND WET SEQUENCES
C
C...PARAMETERS
C       A,B  ARRAYS OF FOURIER SERIES COEFFICIENTS
C       P'S  INITIAL AND TRANSITION PROBABILITES TO BE CALCULATED
C       INITD  DAY OF YEAR ON WHICH THE PROBABILITIES ARE TO START
C       NUM  NUMBER OF DAYS FOR WHICH PROBABILITIES ARE
C         REQUIRED
C       N  ARRAY CONTAINING THE NUMBER OF TERMS FOR THE FOURIER SERIES
C       XB  ARRAY CONTAINING THE MEANS FOR THE F. S..
C
C...DIMENSIONING AND INITIALIZATION
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION A(15,3),B(15,3),P(32),OMP(32),PO1(32),P11(32),
     &P10(32),POO(32),N(3),XB(3)
      COMMON P,OMP,PO1,P11,P10,POO
      LOGICAL*1 LFMT(1)/'*'/
C
C   CALCULATE LAST DAY FOR WHICH PROBABILITY ESTIMATES REQ'D
C
      IEND=INITD+NUM
C
C...INPUT THE NUMBER OF COEFFICIENTS FOLLOWED BY THE MEAN OF
C   THE SERIES AND THE FOURIER COEFFICIENTS
C
      WRITE(6,10)
 10   FORMAT('OINPUT NUMBER OF COEFFICIENTS, IF NUMBER OF ',
     &'COEFFICIENTS = 0, INPUT 1'/1X,'THEN INPUT PAIR OF O S FOR THE'
     &,' COEFFICIENTS'/1X,'OMP, P10, POO')
      READ(5,LFMT) (N(I),I=1,3)
      WRITE(6,30)
 30   FORMAT('OINPUT MEAN FOLLOWED BY PAIRS OF COEFFICIENTS',
     &' ON SEPARATE LINES')
      READ(5,LFMT) (XB(I),I=1,3)
      DO 1 I=1,32
 1    OMP(I)=XB(1)
      DO 2 I=1,32
 2    P10(I)=XB(2)
      DO 3 I=1,32
 3    POO(I)=XB(3)
      DO 20 I=1,3
      IN=N(I)
      DO 40 K=1,IN
 40   READ(5,LFMT) A(K,I),B(K,I)
 20   CONTINUE
C
C...GENERATE A FOURIER SERIES APPROXIMATION TO THE PROBABILITY
C   FOR EACH DAY OF THE YEAR REQUESTED
C
      LB=INITD+1
      LE=IEND+1
      DO 60 LP1=LB,LE
      L=LP1-1
      M=L-INITD+1
C
C...SUM OVER EACH NONZERO HARMONIC TO OBTAIN F.S. ESTIMATE
C
```

```
      IN=N(1)
      DO 70 K=1,IN
      ANG=2.DO*3.141592653589793/365.DO*DFLOAT(K*L)
      DC=DCOS(ANG)
      DS=DSIN(ANG)
70    OMP(M)=OMP(M)+A(K,1)*DC+B(K,1)*DS
      IN=N(2)
      DO 80 K=1,IN
      ANG=2.DO*3.141592653589793/365.DO*DFLOAT(K*L)
      DC=DCOS(ANG)
      DS=DSIN(ANG)
80    P10(M)=P10(M)+A(K,2)*DC+B(K,2)*DS
      IN=N(3)
      DO 90 K=1,IN
      ANG=2.DO*3.141592653589793/365.DO*DFLOAT(K*L)
      DC=DCOS(ANG)
      DS=DSIN(ANG)
90    POO(M)=POO(M)+A(K,3)*DC+B(K,3)*DS
      P(M)=1.DO-OMP(M)
      P11(M)=1.DO-P10(M)
      PO1(M)=1.DO-POO(M)
60    CONTINUE
      RETURN
      END
```

```
      SUBROUTINE DERIV(F,DF,N,H,NSIZ)
C...THIS SUBROUTINE FINDS THE FIRST DERIVATIVE OF EQUISPACED DATA.
C   CENTRAL DIFFERENCE FORMULAS ARE USED FOR ALL POINTS EXCEPT THE
C   FIRST AND THE LAST.  FOR THESE, A QUADRATIC IS PASSED THROUGH
C   THREE SUCCESSIVE POINTS TO OBTAIN THE DERIVATIVE.
C
C...PARAMETERS ARE-
C                  F...THE ARRAY OF FUNCTION VALUES
C                  DF..ARRAY OF DERIVATIVE VALUES
C                  N...THE NUMBER OF POINTS
C                  H...THE UNIFORM SPACING BETWEEN X VALUES
C
C...FROM C. F. GERALD, APPLIED NUMERICAL ANALYSIS, 2ND ED.
C    1978 BY ADDISON WESLEY
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION F(NSIZ),DF(NSIZ)
C...COMPUTE THE DERIVATIVES AT X(2) THROUGH X(N-1)
      NM1=N-1
      DO 10 I=2,NM1
        DF(I)=(F(I+1)-F(I-1))/2.DO/H
 10   CONTINUE
C...NOW COMPUTE DERIVATIVE AT X(1)
      DF(1)=(2.DO*F(2)-1.5DO*F(1)-0.5DO*F(3))/H
C...AND GET IT AT X(N)
      DF(N)=(1.5DO*F(N)-2.DO*F(N-1)+0.5DO*F(N-2))/H
      RETURN
      END
```

```
      SUBROUTINE SIMPS(F,N,H,RESULT)
C
C...THIS ROUTINE FROM C. F. GERALD, 1978
C
C...THIS ROUTINE PERFORMS SIMPSON'S RULE INTEGRATION OF A FUNCTION
C   DEFINED BY A TABLE OF EQUISPACED VALUES.
C...THE ROUTINE HAS BEEN MODIFIED TO USE THE TRAPEZOIDAL
C   RULE FOR 1 PANEL AND GIVEN AN EXIT POINT WHEN ONLY 3 PANELS
C   ARE ENCOUNTERED.
C...PARAMETERS ARE -
C     F      ARRAY OF VALUES OF THE FUNCTION
C     N      NUMBER OF POINTS
C     H      UNIFORM SPACING BETWEEN X VALUES
C     RESULT ESTIMATE OF THE INTEGRAL THAT IS RETURNED TO CALLER
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION F(1000)
C...CHECK TO SEE IF NUMBER OF PANELS IS EVEN.  NUMBER OF PANELS
C   IS N-1.
      NPANEL=N-1
      NHALF=NPANEL/2
      NBEGIN=1
      RESULT=0.D0
      IF((NPANEL-2*NHALF).EQ.0) GO TO 5
      IF(NPANEL.EQ.1) GO TO 15
C...NUMBER OF PANELS IS ODD. USE 3/8 RULE OF FIRST THREE
C   PANELS, 1/3 RULE ON REST OF THEM.
      RESULT=3.D0*H/8.D0*(F(1)+3.D0*F(2)+3.D0*F(3)+F(4))
      NBEGIN=4
      IF(NBEGIN.EQ.N) RETURN
C...APPLY 1/3 RULE - ADD IN FIRST,SECOND, LAST VALUES
 5    RESULT=RESULT+H/3.D0*(F(NBEGIN)+4.D0*F(NBEGIN+1)+F(N))
      NBEGIN=NBEGIN+2
      IF(NBEGIN.EQ.N) RETURN
C...THE PATTERN AFTER NBEGIN+2 IS REPETITIVE. GET NEND, THE
C   PLACE TO STOP.
      NEND=N-2
      DO 10 I=NBEGIN,NEND,2
 10   RESULT=RESULT+H/3.D0*(2.D0*F(I)+4.D0*F(I+1))
      RETURN
C
C...FOR NPANEL EQUAL TO 1 USE THE TRAPEZOIDAL RULE
C
 15   RESULT=(F(1)+F(2))*H/2.D0
      RETURN
      END
```

```
C                                   GAM2 (WONG, 1980)
C
C
C LIKELIHOOD RATIO TEST FOR GAMMA SCALE DIFFERENCE WITH COMMON SHAPE
C ITERATION BY SEQUENTIAL SUBSTITUTION
C REFERENCE PAUL W MIELKE(1976) JAM V15 NO 2 &181-183
C    I/O  5...INPUT PARAMETERS AND TITLES
C         6...OUTPUT
C         7...SAMPLE SIZE AND SAMPLE 1 DATA
C         8...SAMPLE SIZE ANO SAMPLE 2 DATA
C
C     IF JOPT1=1, OUTPUT SHAPE PARAMETERS AS THEY CONVERGE
C     XLT=0.0001 (CONVERGENCE CRITERIA FOR SHAPE PARAMETER)
C     MLX=MAXIMUM NUMBER OF ITERATIONS PERMITTED (USUALLY 100)
C     TITLE1(I)=ANY TITLE PERTAINING TO SAMPLE ONE
C     N1=SIZE OF SAMPLE 1
C     TITLE2(I)=ANY TITLE PERTAINING TO SAMPLE TWO
C     N2=SIZE OF SAMPLE 2
C     OUTPUT :
C     NUMBER OF OBSERVTIONS IN EACH SAMPLE, THE CONVERGENCE CRITERION,
C     CONVERGENCE OF THE SHAPE PARAMETER (OPTIONAL), NUMBER OF STEPS FOR
C     CONVERGENCE, ANO TABLE OF SUMMARY STATISTICS, ESTIMATES OF
C
C     PARAMETERS UNDER THE TWO HYPOTHESES (ALPHA1=ALPHA2=ALPHA,
C     BETA1=BETA2=BETA) ANO (ALPHA1=ALPHA2=ALPHA,BETA1,BETA2)
C     AND THE CHI-SQUARE VALUE.
C     PARAMETERS WHICH MUST BE INPUT IN ORDER TO EXECUTE PROGRAM :
C     LINE 1 :JOPT1
C     LINE 2 :XLT,MLX
C     LINE 3 :TITLE1(I),I=1,10   (10A8)
C     LINE 4 :N1
C     LINE 5 :DATA FROM SAMPLE ONE (ONE PER LINE)
C        .
C        .
C        .
C     LINE N1+5 :TITLE2(I),I=1,10   (10A8)
C     LINE N1+6 :N2
C     LINE N1+7 :DATA FROM SAMPLE TWO (ONE PER LINE)
C
C...INITIALIZATION
C
      IMPLICIT REAL*8(A-H,O-Z)
      LOGICAL*1 LFMT(1)/'*'/
      DIMENSION X1(1000),X2(1000)
      REAL*8 TITLE1(10)
      REAL*8 TITLE2(10)
      COMMON C,XLT,NS,MLX
    7 FORMAT('1')
    8 FORMAT('0','THE SHAPE PARAMETER AS IT CONVERGES',//,
     -'INITIAL VALUE:',1X,F12.4)
   11 FORMAT('1',///,'GAMMA LIKELIHOOD RATIO TEST',///,
     -'SAMPLE 1 :-',3X,10A8,/,15X,
     -'THE NUMBER OF OBSERVATIONS IS',I4,//,
     -'VS',//,'SAMPLE 2 :-',3X,10A8,/,15X,
     -'THE NUMBER OF OBSERVATIONS IS',I4,
     -///,'THE CONVERGENCE CRITERION IS   ',F12.10)
   12 FORMAT(10A8)
   17 FORMAT(' ',////,'FOR THE POOLED SAMPLE ')
   18 FORMAT(' ',///,'FOR THE INDIVIDUAL SAMPLES')
   22 FORMAT('0','FOR THE HYPOTHESIS A1=A2=A,B1=B2=B')
   23 FORMAT('0',5X,'THE GAMMA  SCALE PARAMETER IS ',F10.4)
   24 FORMAT(' ',5X,'THE GAMMA  SHAPE PARAMETER IS ',F10.4)
   25 FORMAT('0','THE LOG OF THE LIKELIHOOD FUNCTION IS ',F10.4)
   26 FORMAT('-','FOR THE HYPOTHESIS A1=A2=A,B1,B2')
```

```
   27 FDRMAT('O',5X,'THE CDMMON SHAPE PARAMETER IS ',F10.4)
   28 FDRMAT(' ',5X,'THE  SCALE  FDR  SAMPLE 1  IS ',F10.4)
   29 FORMAT(' ',5X,'THE  SCALE  FOR  SAMPLE 2  IS ',F10.4)
   30 FDRMAT('-','THE CHI-SQUARE VALUE IS ',F10.4,///)
   31 FORMAT(' ',///,'TABLE DF SUMMARY STATISTICS')
      C=0.577215665
      NS=25
C
C...READ IN PARAMETERS AND DATA
C
      READ(5,LFMT) JDPT1
      READ(5,LFMT) XLT,MLX
      READ(5,12)(TITLE1(I),I=1,10)
      READ(7,LFMT) N1
      DD 200 I=1,N1
 200  READ(7,LFMT) X1(I)
      READ(5,12)(TITLE2(I),I=1,10)
      READ(8,LFMT) N2
      DO 210 I=1,N2
 210  READ(8,LFMT) X2(I)
      WRITE(6,11)(TITLE1(I),I=1,10),N1,(TITLE2(I),I=1,10),N2,XLT
C
C...CALCULATE TDTALS AND MEANS
C
      SUM1=0.0
      SUM2=0.0
      SUM3=0.0
      SUM4=0.0
      DO 15 I=1,N1
      SUM1=SUM1+X1(I)
      SUM2=SUM2+DLDG(X1(I))
   15 CDNTINUE
      DD 16 I=1,N2
      SUM3=SUM3+X2(I)
      SUM4=SUM4+DLDG(X2(I))
   16 CDNTINUE
      XM1=SUM1/N1
      XM2=SUM3/N2
      XL1=SUM2/N1
      XL2=SUM4/N2
      AAA=(N1*DLDG(XM1)+N2*DLDG(XM2)-SUM2-SUM4)/(N1+N2)
      XMP=(SUM1+SUM3)/(N1+N2)
      XLP=(SUM2+SUM4)/(N1+N2)
      AA=DLOG(XMP)-XLP
      AP=1.0
      WRITE(6,17)
      IF (JDPT1.EQ.1) WRITE(6,8)AP
C
C...CALCULATE SHAPE PARAMETER FDR THE PDOLED SAMPLES
C
      CALL ITS(AP,AA,JOPT1)
      A12=AP
      WRITE(6,18)
      IF (JDPT1.EQ.1) WRITE(6,8)AP
C
C...CALCULATE SHAPE PARAMETER FOR INDIVIDUAL SAMPLES
C
      CALL ITS(A12,AAA,JDPT1)
C
C...CALCULATE SCALE PARAMETERS...NDTE THIS ROUTINE WDRKS WITH
C   THE INVERSE DF THE SCALE PARAMETERS USED ELSWHERE IN THE THESIS
C
      BP=XMP/AP
      B1=XM1/A12
      B2=XM2/A12
      IF (JDPT1.EQ.1) WRITE(6,7)
```

```
      WRITE(6,31)
      WRITE(6,22)
      WRITE(6,23) BP
      WRITE(6,24) AP
C
C...CALCULATE MAX LIKELIHOOD STATISTIC
C
      T1=-AP*(N1+N2)*DLOG(BP)-(N1+N2)*DLGAMA(AP)+(AP-1.0)*(SUM2+SUM4)
     1-(SUM1+SUM3)/BP
      T2=-N1*A12*DLOG(B1)-N1*DLGAMA(A12)+(A12-1.0)*SUM2-SUM1/B1-N2*A12
     1*DLOG(B2)-N2*DLGAMA(A12)+(A12-1.0)*SUM4-SUM3/B2
      TEST=2.0*(T2-T1)
      WRITE(6,25) T1
      WRITE(6,26)
      WRITE(6,27) A12
      WRITE(6,28) B1
      WRITE(6,29) B2
      WRITE(6,25) T2
      WRITE(6,30) TEST
      STOP
      END
```

```
      SUBROUTINE ITS(AL,AX,JOPT1)
C
C...THIS SUBROUTINE CALCULATES THE SHAPE PARAMETER USING MIELKE'S
C   ITERATIVE PROCEDURE (MIELKE, 1976)
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON C,XLT,NS,MLX
      II=0
  101 AO=AL
      II=II+1
      IF(II.GE.MLX)GO TO 105
      XA=(AL*(NS+0.5))/(NS+AL-0.5)
      XSUM=0.0
      DO 102 I=1,NS
      XSUM=XSUM+1.0/(I*(I+AL-1.0))
  102 CONTINUE
      AL=1.0+(DLOG(XA)+C-AX)/XSUM
      IF (JOPT1.EQ.1) WRITE(6,100)AL
  100 FORMAT(' ',15X,F12.4)
      IF(DABS(AO-AL).LT.XLT)GO TO 103
      GO TO 101
  103 WRITE(6,104) II
  104 FORMAT('0','CONVERGENCE IN ',I3,' STEPS')
      GO TO 107
  105 WRITE(6,106) II
  106 FORMAT(' ',' NO CONVERGENCE IN ',I3,' STEPS')
  107 RETURN
      END
```